The Dissertation Committee for Jun Zhang
certifies that this is the approved version of the following dissertation:

Adapting MIMO Networks to Manage Interference

Committee:

Jeffrey G. Andrews, Supervisor

Robert W. Heath, Jr.

Sriram Vishwanath

Arunabha Ghosh

Lili Qiu

Adapting MIMO Networks to Manage Interference

by

Jun Zhang, B.Eng.; M.Phil.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

December 2009

To my family.

Acknowledgments

First of all, I would like to thank my supervisor Prof. Jeffrey Andrews, for recruiting me to WNCG, for providing the comfortable research environment, for his invaluable guidance and never-ending encouragement. He is the best supervisor a graduate student can ever hope for. I would like to thank my committee, Prof. Robert Heath, Prof. Sriram Vishwanath, Dr. Arunabha Ghosh, and Prof. Lili Qiu, for their helpful research advice and constructive feedback on my dissertation. I especially thank Prof. Heath, who inspired my research on MIMO communications and provided invaluable advices during my research. I thank AT&T Labs for graciously supporting my research, and thank Dr. Arunabha Ghosh and David Wolter for inviting me for summer internships at AT&T Labs during 2007-08.

I take this opportunity to express my appreciation to my colleagues at WNCG, especially Kaibin Huang, Runhua Chen, Vikram Chandrasekhar, Marios Kountouris, and Andrew Hunter, for their helpful discussions and friendships. I also thank Janet Preuss of WNCG for taking care of various paperwork, which greatly facilitates my study and research.

Most important of all, I would like to thank my parents, Nanxian Zhang and Zhiying Miao, for their unconditional love and support throughout my life, especially in the past few years when I was pursuing my PhD overseas. I would also like to thank my sisters, Ying and Li, for their support and encouragement.

Adapting MIMO Networks to Manage Interference

Publication No. _____

Jun Zhang, Ph.D.
The University of Texas at Austin, 2009

Supervisor: Jeffrey G. Andrews

Multiple-Input Multiple-Output (MIMO) communication uses multiple transmit and receive antennas to improve the throughput in wireless channels. In cellular networks, self-interference greatly degrades MIMO's potential gain, especially in multiuser MIMO systems where multiple users in each cell share the spatial channel in order to maximize the total throughput. In a multiuser MIMO downlink, the two main causes of this self-interference are residual inter-user interference due to imperfect spatial separation between the users and other-cell interference due to cochannel transmissions in other cells. This dissertation develops adaptive transmission strategies to deal with both residual inter-user interference and other-cell interference in cellular MIMO networks.

For the residual inter-user interference caused by imperfect channel state information at the transmitter, we explicitly characterize the impact of channel quantization and feedback delay. Achievable ergodic rates for both single-user and multiuser MIMO systems with different channel state information are derived. Adaptive switching between single-user and multiuser MIMO modes is proposed to improve the throughput, based on the accuracy of the available channel information. It is then extended to a multi-mode transmission strategy which adaptively adjusts the number of active users to control residual interference and provide additional array gain.

To adaptively minimize the other-cell interference, two practical base station coordination strategies are proposed. The first is a cluster based coordination algorithm with different coordination strategies for cluster interior and cluster edge

users. It performs full intra-cluster coordination for enhancing the sum throughput and limited inter-cluster coordination for reducing the interference for cluster edge users. A multi-cell linear precoder is designed to perform the coordination. The second is an adaptive intercell interference cancellation strategy, where multiple base stations jointly select transmission techniques based on user locations to maximize the sum throughput. Spatial interference cancellation is applied to suppress other-cell interference. Closed-form expressions are derived for the achievable throughput, and the proposed adaptive strategy is shown to provide significant average and edge throughput gain. The feedback design to assist the interference cancellation is also discussed.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

The wireless industry is one of the fastest growing businesses in the world, led by the cellular systems. It was only in late 1983 that the first commercial cellular telephone system in the US, called Advanced Mobile Phone Service (AMPS), was deployed in the Chicago area. By the end of 2008, there were over 4 billion mobile subscribers in the world, which is more than double the number of fixed line subscribers and amounts to a penetration rate higher than 60 percent [71]. Wireless communications, by eventually providing seamless connectivity, will continue to bring revolutionary changes to the way people communicate, work, and entertain. Different wireless communities have set ambitious targets to meet the growing demand for high data rates during their standardization work, e.g. the peak downlink data rate is 100 Mbps for 3GPP-LTE (Long Term Evolution) developed by 3rd Generation Partnership Project (3GPP) [1], while the ITU (International Telecommunication Union) definition of a 4G system, called IMT-Advanced, requires a target downlink peak data rate of 1 Gbps [2]. To meet the service and performance requirements, advanced radio technologies need to be employed. Multi-antenna transmission and reception, known as Multiple-Input Multiple-Output (MIMO), has proven to be one of the key techniques for broadband wireless networks.

MIMO systems employ multiple antennas at the transmitter and/or receiver to improve the transmission throughput and reliability over wireless channels. Since the theoretical breakthrough in the late 1990's [46, 136], MIMO communications have been extensively researched [15, 54, 57, 112] and have transitioned from a theoretical concept to a key technique in commercial wireless standards, including 3GPP-LTE [95], 802.16/WiMAX [11], and IEEE 802.11n [155]. However, the current applications of MIMO mainly focus on the point-to-point link, i.e. applying the single-user MIMO

Table 1.1: Supported MIMO Techniques in Different Standards

	SU-MIMO	MU-MIMO	Networked MIMO
IEEE 802.11n	√		
IEEE 802.16e	√		
3GPP LTE	√	√	
3GPP LTE-Advanced	√	√	√

technique. To fully exploit the benefits of MIMO communications, the application in the network setting should be considered. This dissertation targets the cellular MIMO network, and tries to exploit MIMO's capacity gain by efficiently suppressing the existing self-interference.

We start this introduction chapter with a brief description of different MIMO communication systems. Then we point out the challenges of self-interference in cellular MIMO networks – residual inter-user interference and other-cell interference, which are the main motivations of this dissertation. Finally, a summary of contributions in this dissertation is provided, followed by the organization of the following chapters.

1.1 The Key Types of MIMO Communication Systems

Based on the number of nodes at the transmit and receive ends, MIMO communication systems can be divided into single-user MIMO (SU-MIMO), multiuser MIMO (MU-MIMO), and networked MIMO, as shown in Figure 1.1. The supported MIMO techniques in different broadband wireless standards are listed in Table 1.1. Note that IEEE 802.11n, IEEE 802.16e, and 3GPP LTE have been finalized, while 3GPP LTE-Advanced is under development. From this table, we see that so far only SU-MIMO has been widely implemented in commercial systems. In this section, we show the advantages of MIMO communications, and also indicate the main difficulties faced in practice.

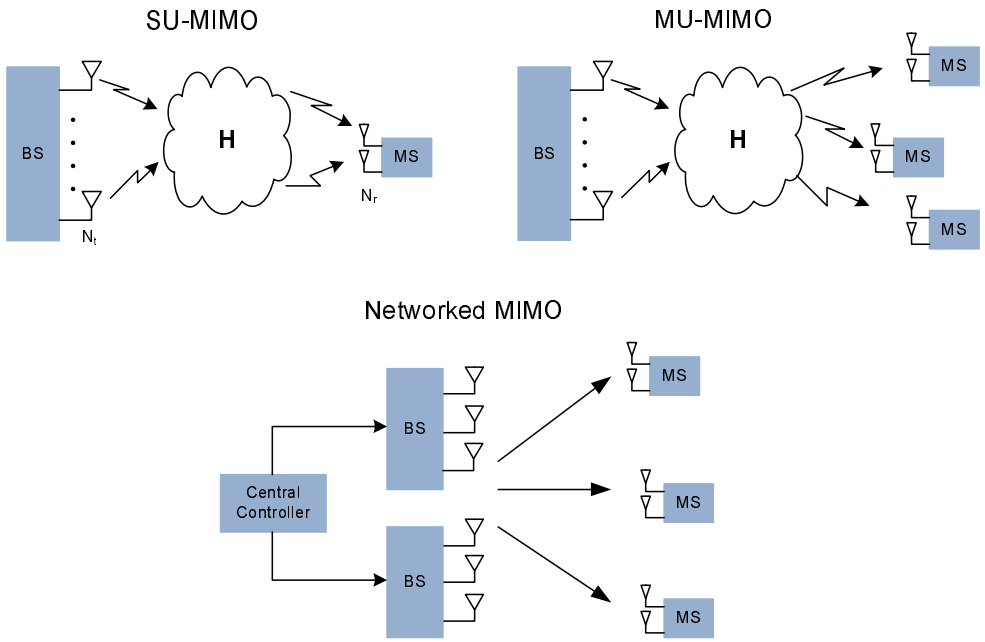


Figure 1.1: Different MIMO systems in the cellular network, where "BS" denotes the base station and "MS" denotes the mobile station.

1.1.1 Single-user MIMO

For a point-to-point link, i.e. an SU-MIMO system, with N_t transmit antennas and N_r receive antennas, it was shown in the late 1990's that the capacity grows linearly with $\min(N_t, N_r)$ in Rayleigh fading with coherent reception at high SNR [46, 136]. Later, it has been shown that this behavior holds under rather general channel conditions, e.g. with heterogeneous arrays or channel correlation [34, 139], and in non-coherent channels where neither the transmitter nor the receiver knows the channel realization [170]. This throughput gain is achieved by sending multiple independent data streams simultaneously through multiple spatial channels, which is called *spatial multiplexing*. As the capacity gain is obtained without expanding the communication bandwidth, it makes MIMO a desirable technique to provide high data rate for bandwidth-constrained wireless networks. Besides spatial multiplexing, multiple antennas can also be used to provide *spatial diversity gain* through receive combining or space-time coding [9, 134, 135]. Spatial diversity can be used to im-

prove the transmission reliability over the fading channel and increase the coverage in cellular networks. There is a fundamental tradeoff between the achievable spatial multiplexing gain and the spatial diversity gain [149, 171].

Due to the significant performance gain, both spatial multiplexing and spatial diversity have been included in current and future commercial wireless standards, as shown in Table 1.1. In practical systems, the MIMO technique can be classified into closed-loop (CL) MIMO modes and open-loop (OL) MIMO modes. The OL MIMO modes require no or little feedback from mobile users, so it is suitable for scenarios where accurate feedback is difficult to get or the channel changes too fast, such as the high mobility scenario. On the other hand, the CL MIMO modes require explicit feedback, and are able to provide additional performance gain and simplify the receiver structure. The precoder of CL MIMO modes is usually based on the predefined codebook, and the index of the selected codeword is fed back to the transmitter. This technique is called *limited feedback* [97], and it has been successfully deployed in commercial systems. For example, 3GPP LTE employs OL MIMO modes including transmit diversity and OL spatial multiplexing, and CL MIMO modes including codebook based precoding and spatial multiplexing [3].

1.1.2 Multiuser MIMO

Due to space constraints, mobile terminals can only have a small number of antennas, normally one or two, which bounds the capacity gain of SU-MIMO in cellular networks. MU-MIMO provides the opportunity to overcome such a limitation by communicating with multiple mobiles simultaneously. It effectively increases the dimension of the equivalent spatial channel and provides spatial multiplexing gain proportional to the number of transmit antennas at the base station (BS) even with single-antenna mobiles. The uplink/multiaccess channel of MU-MIMO is a straightforward extension of the point-to-point MIMO link [136]. For the downlink/broadcast channel, it has been shown that dirty-paper coding (DPC) [36] achieves the capacity [25, 145, 146, 151, 157]. Channel state information at the transmitter (CSIT) is required for MIMO downlink, which is used to separate signals for different users

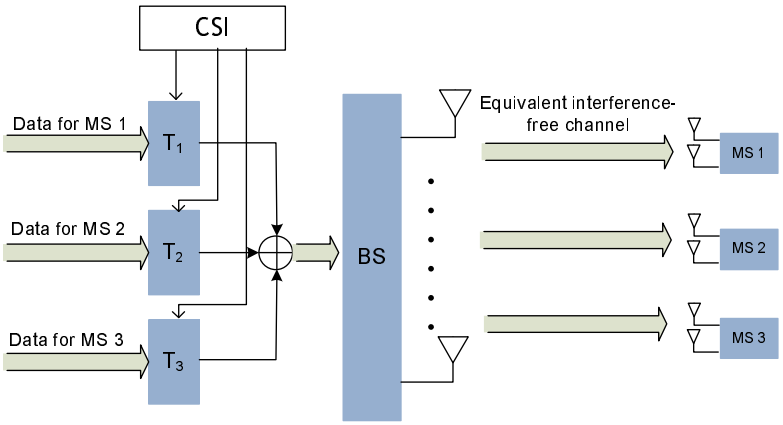


Figure 1.2: The MU-MIMO system with zero-forcing linear precoding.

in the spatial domain. In addition, MU-MIMO has higher immunity to propagation limitations faced by SU-MIMO, such as channel rank loss and antenna correlation [53].

DPC, while theoretically optimal, is an information theoretic concept that is difficult to implement in practice. Linear precoding for MU-MIMO is a more practical option, which suppresses inter-user interference by applying a linear transformation of the transmitted signal vector. Figure 1.2 shows the MIMO downlink with linear precoding based on the zero-forcing criterion, which provides each user an interference-free channel with properly designed precoding matrices. *Zero-forcing (ZF)* precoding and *block diagonalization (BD)* are two important MU-MIMO linear precoding schemes [30, 32, 110, 128, 133]. Both are designed to completely precancel inter-user interference at the transmitter. The difference is that ZF has a diagonal effective channel whereas BD has a block diagonal effective channel for each user. For the MU-MIMO systems explored in this dissertation, we mainly apply ZF and BD precoding, as they are low complexity, easy to analyze, and also provide near-to-optimal performance [74, 125, 127]. To further improve the performance, other precoding techniques can be employed, such as coordinated beamforming [28, 29, 133], which jointly designs the precoder and decoder.

As shown in Table 1.1, there is limited support for MU-MIMO in the current

commercial wireless standards. A major difficulty is the requirement of CSIT, which will be discussed later in this chapter. Only 3GPP LTE supports MU-MIMO, where the MU-MIMO precoder in the downlink reuses the SU-MIMO precoder and no inter-user interference cancelation is applied during the precoding process. This kind of MU-MIMO mode is able to provide throughput gain in correlated channels [116]. It is expected that LTE-Advanced will have more aggressive inclusion of MU-MIMO.

1.1.3 Networked MIMO

An important extension of MIMO techniques for cellular networks is *networked MIMO*, also called *base station coordination* or *multi-cell processing* [122]. In networked MIMO, multiple BSs collaborate with each other to communicate with multiple users. In this way, these BSs can be collectively regarded as a single "super BS", and the whole system is equivalent to a virtual MIMO system. In theory, with enough cooperation between the BSs, networked MIMO can eliminate other-cell interference completely and therefore change the cellular network from an interference-limited system to a noise-limited system [73, 132]. Thus, networked MIMO has drawn significant attention, and is regarded as a key enabling technique for 3GPP LTE-Advanced, where it is called *CoMP (coordinated multiple point transmission and reception)*. In the uplink, the super BS can separate signals for different users with multiuser detection. For the downlink, assuming the availability of channel state information and traffic data of all users, multiple base stations can jointly design downlink precoders to pre-cancel inter-user interference for different users. Both the optimal DPC and suboptimal linear precoding have been investigated for networked MIMO [48, 79, 83, 123, 131, 159].

As in MU-MIMO systems, there are practical constraints that limit the most powerful forms of networked MIMO. One of the major difficulties is the requirement of timely channel state information at multiple base stations, the amount of which is proportional to both the number of base station antennas and the number of base stations. This poses challenges for channel estimation, feedback, and inter-cell communication (backhaul), which generally has its limitations both in terms of rate

and delay. In addition, the complexity of optimal processing is quite high, and grows rapidly with the network size. In this dissertation, we will develop practical strategies to deal with some of these issues.

1.2 Challenges in Cellular MIMO Networks

In cellular networks, self-interference greatly degrades MIMO's potential gain. In this dissertation, we focus on the MIMO downlink, and consider two main causes of this self-interference: residual same-cell inter-user interference due to imperfect spatial separation between different users, and other-cell interference due to cochannel transmissions in other cells. In this section, we demonstrate the challenges brought by the self-interference in cellular MIMO networks and show the motivation of our research.

1.2.1 Residual Inter-user Interference

As shown in Table 1.1, there is limited support for MU-MIMO in current wireless standards. This is because there are many technical challenges that must be overcome to exploit the full benefits of MU-MIMO. A major one is the requirement of channel state information at the transmitter (CSIT), which is difficult to get especially with non-trivial mobility. For the MIMO broadcast channel with N_t transmit antennas and N_r receive antennas, with full CSIT the sum throughput can grow linearly with N_t even when $N_r = 1$, but without CSIT the spatial multiplexing gain is the same as for SU-MIMO, i.e. the throughput grows linearly with $\min(N_t, N_r)$ at high SNR [61]. Therefore, CSIT plays a significant role in MU-MIMO in cellular networks, where N_r is typically very small.

Limited feedback has proved an efficient technique to provide partial CSI to the transmitter for point-to-point MIMO link [98, 99, 103, 104]. Recently, limited feedback for MU-MIMO has been researched, which provides quantized channel information to the transmitter [21, 43, 75, 114, 156]. However, with such imperfect channel state information, there will be residual inter-user interference due to imperfect spatial

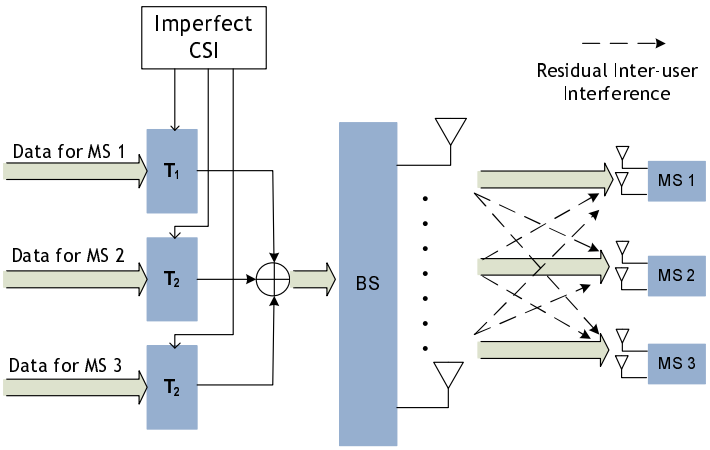


Figure 1.3: The MU-MIMO system with imperfect CSI.

separation between different users, as shown in Figure 1.3. To achieve the full spatial multiplexing gain, the number of feedback bits needs to increase linearly with both SNR (in dB) and the number of transmit antennas [75, 114], which is normally not possible. Besides quantization, there are other imperfections in the available CSIT, such as estimation error and feedback delay. Cumulatively, such imperfections will greatly degrade the performance of MU-MIMO [92]. In this dissertation, we will investigate the impact of imperfect CSIT on the performance of MU-MIMO systems, and propose adaptive strategies to improve the throughput.

1.2.2 Other-cell Interference

Another important cause of self-interference in cellular MIMO networks is other-cell interference (OCI), as shown in Figure 1.4. In a commercial cellular system the transmission in each cell acts as interference to other cells, and the entire network is essentially interference-limited by design. While the problem of OCI is inherent to cellular systems, its impact on MIMO transmission is more significant because each neighboring BS antenna element can act as a unique interfering source, thereby making it difficult for the mobile to estimate and suppress them [12]. With N_r receive antennas, each mobile can only cancel/decode up to N_r different sources using linear techniques [112]. Furthermore, interference is more severe for the downlink because

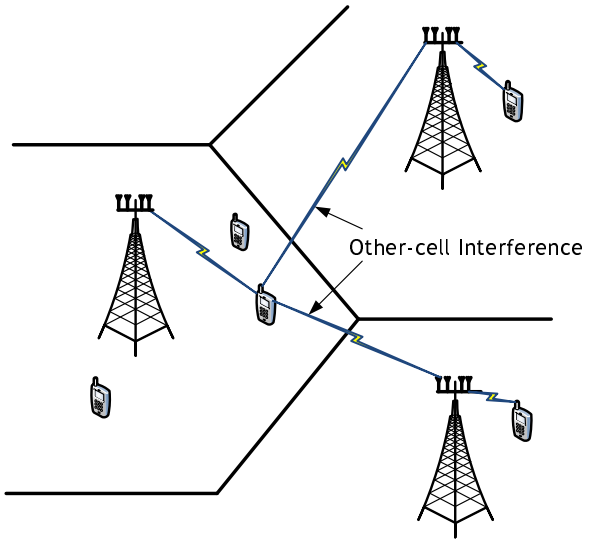


Figure 1.4: Other-cell interference in cellular networks.

complicated interference suppression techniques are not practical for mobile terminals, which need to be power-efficient and compact. Coordination between users is usually not allowed. The capacity gains promised by MIMO techniques have been shown to degrade severely in the multicell environment [16, 27, 38]. Conventional approaches to mitigate multicell interference, such as static frequency reuse, sectoring, and spread spectrum, are not efficient for MIMO networks as each has important drawbacks [12]. The difficulty in combating interference for MIMO is essentially due to the limitation of spatial degrees of freedom, most of which are used to suppress the spatial interference introduced by spatial multiplexing at the cell site while few are left to suppress other-cell interference.

Thanks to the fast improvement of processing capability at BSs and the increase of the backhaul capacity, coordinated multicell MIMO communications with cooperative processing among BSs have drawn a significant amount of interest in recent years. The conventional MIMO network with single-cell processing forms a MIMO interference channel, whose spatial degrees of freedom are determined by the number of transmit antennas at each BS [72]. With full coordination across B BSs and a large number of mobile users, the coordination system forms a virtual MIMO

broadcast channel, which increases the spatial degrees of freedom by B times. The BS coordination has been shown to be an efficient way to combat OCI and provide great advantages over single-BS processing [83, 123, 159]. The full BS coordination, however, is of very high complexity and requires a large amount of information exchange. This dissertation will investigate practical BS coordination strategies to suppress OCI for cellular MIMO networks.

1.3 Overview of Contributions

This dissertation focuses on the MIMO downlink in cellular networks. The main purpose is to investigate the impact of self-interference and propose adaptive strategies to improve the downlink spectral efficiency. Following are the main contributions of this dissertation:

1. **SU-MIMO vs MU-MIMO with Imperfect CSIT:** We compare the throughput of both SU-MIMO and MU-MIMO in the broadcast channel with imperfect CSIT due to delay and channel quantization. Accurate closed-form approximations are derived for achievable rates for both SU and MU-MIMO. It is shown that SU-MIMO is relatively robust to delayed and quantized channel information, while MU-MIMO with zero-forcing precoding loses its spatial multiplexing gain with a fixed delay or fixed codebook size. Based on derived achievable rates, a mode switching algorithm is proposed, which switches between SU and MU-MIMO modes to improve the spectral efficiency based on average SNR, normalized Doppler frequency, and the channel quantization codebook size. The operating regions for SU and MU modes are determined, and it is shown that the MU mode is active only when the normalized Doppler frequency is very small and the codebook size is large.

2. **Multi-mode Transmission:** To further improve the spectral efficiency achieved by the MIMO broadcast channel with delayed and quantized channel state information, a multi-mode transmission strategy is proposed in Chapter 3. It adaptively

adjusts the number of active users, denoted as the *transmission mode*, to balance transmit array gain, spatial division multiplexing gain, and residual inter-user interference. Accurate closed-form approximations are derived for the achievable rates for different modes, which are used to select the active mode that maximizes the ergodic throughput. User scheduling algorithms based on multi-mode transmission are then proposed for the network with a large number of users, to reduce the overall amount of feedback. It is shown that the proposed algorithms provide throughput gains at moderate yet practically relevant SNR.

3. **Clustered BS Coordination:** To efficiently suppress OCI with BS coordination while keep the overhead and complexity low, a clustered coordination strategy is proposed, which includes full intra-cluster coordination–to enhance the sum rate– and limited inter-cluster coordination–to reduce interference for the cluster edge users. Multi-cell block diagonalization is used to coordinate the transmissions across multiple BSs in the same cluster. To satisfy per-BS power constraints, three combined precoder and power allocation algorithms are proposed with different performance and complexity tradeoffs. For inter-cluster coordination, the coordination area is chosen to balance fairness for edge users and the achievable sum rate. It is shown that a small cluster size (about 7 cells) is sufficient to obtain most of the sum rate benefits from clustered coordination while greatly relieving channel feedback requirement. Simulations show that the proposed coordination strategy efficiently reduces interference and provides a considerable sum rate gain for cellular MIMO networks.

4. **Adaptive Intercell Interference Cancellation:** In Chapter 5, another BS coordination strategy is proposed, called adaptive spatial intercell interference cancellation (ICIC). OCI suppression is achieved by zero-forcing precoding at the BS. A principle question we explore is whether it is better to do ICIC or simply standard single-cell beamforming. We explore this question analytically and show that beamforming is preferred for all users when the edge SNR is low (< 0 dB), and ICIC is preferred when the edge SNR is high (> 10 dB), for example in an urban

setting. At medium SNR, a proposed adaptive strategy, where multiple base stations jointly select transmission strategies based on the user location, outperforms both while requiring a lower feedback rate than the pure ICIC approach. The employed metric is sum rate, which is normally a dubious metric for cellular systems, but surprisingly we show that even with this reward function the adaptive strategy also improves fairness. When the channel information is provided by limited feedback, the impact of the induced quantization error is also investigated. It is shown that ICIC with well-designed feedback strategies still provides significant throughput gain.

1.4 Organization

The remainder of this dissertation is organized as follows. Chapter 2 compares SU-MIMO and MU-MIMO in the MIMO broadcast channel with imperfect CSIT, and proposes an SU/MU mode switching algorithm to improve the spectral efficiency. Chapter 3 proposes a multi-mode transmission strategy to further improve the throughput, which adjusts the number of active users to balance transmit array gain, spatial division multiplexing gain, and residual inter-user interference. In Chapter 4, a clustered BS coordination strategy with linear precoding is proposed to suppress other-cell interference, while Chapter 5 proposes an adaptive spatial intercell interference cancellation strategy. Finally, conclusions are made in Chapter 6.

Chapter 2

SU-MIMO vs MU-MIMO with Imperfect CSIT

Channel state information at the transmitter (CSIT) is required for MU-MIMO in the downlink. Limited feedback is an efficient way to provide partial CSIT, which feeds back the quantized channel information to the transmitter via a low-rate feedback channel [97, 98]. Besides quantization, there are other imperfections in the available CSIT, such as estimation error and feedback delay. Imperfect CSIT degrades the performance of MIMO communications, and its effects on SU-MIMO and MU-MIMO transmissions are quite different: MU-MIMO suffers from residual inter-user interference which limits the achievable throughput at high SNR while SU-MIMO only suffers a power loss. With imperfect CSIT, it is not clear whether–or more to the point, when– MU-MIMO can outperform SU-MIMO. In this chapter, we are interested in the following question: *With imperfect CSIT, including delay and channel quantization, when can MU-MIMO actually deliver a throughput gain over SU-MIMO?*

2.1 Related Work

For the MIMO broadcast channel, CSIT is required to separate the spatial channels for different users. To obtain the full spatial multiplexing gain for MU-MIMO systems employing zero-forcing (ZF) or block-diagonalization (BD) precoding, it was shown in [75, 114] that the quantization codebook size for limited feedback needs to increase linearly with SNR (in dB) and the number of transmit antennas. Zero-forcing dirty-paper coding and channel inversion systems with limited feedback were investigated in [43], where a sum rate ceiling due to a fixed codebook size was derived for both schemes. In [156], it was shown that to exploit multiuser diversity for ZF, both channel direction and information about signal-to-interference-plus-noise

ratio (SINR) must be fed back. In [20], it was shown that the feedback delay limits the performance of joint precoding and scheduling schemes for the MIMO broadcast channel at moderate levels of Doppler. More recently, a comprehensive study of the MIMO broadcast channel with ZF precoding was done in [21], which considered downlink training and explicit channel feedback and concluded that significant downlink throughput is achievable with efficient CSI feedback. For a compound MIMO broadcast channel, the information theoretic analysis in [23] showed that scaling the CSIT quality such that the CSIT error is dominated by the inverse of SNR is both necessary and sufficient to achieve the full spatial multiplexing gain.

Although previous studies show that the spatial multiplexing gain of MU-MIMO can be achieved with limited feedback, it requires the codebook size to increase with SNR and the number of transmit antennas. Even if such a requirement is satisfied, there is an inevitable rate loss due to quantization error, plus other CSIT imperfections such as estimation error and delay. In addition, most of prior work focused on the achievable spatial multiplexing gain, mainly based on the analysis of the rate loss due to imperfect CSIT, which is usually a loose bound [23, 75, 114]. Such analysis cannot accurately characterize the throughput loss, and no comparison with SU-MIMO has been made.

There are several related studies comparing space division multiple access (SDMA) and time division multiple access (TDMA) in the multi-antenna broadcast channel with limited feedback and with a large number of users. TDMA and SDMA with different scalar feedback schemes for scheduling were compared in [41], which shows that SDMA outperforms TDMA as the number of users becomes large while TDMA outperforms SDMA at high SNR. TDMA and SDMA with opportunistic beamforming were compared in [80], which proposed to adapt the number of beams to the number of active users to improve the throughput. A distributed mode selection algorithm switching between TDMA and SDMA was proposed in [91], where each user feeds back its preferred mode and the channel quality information.

2.2 Contributions

In this chapter, we compare SU-MIMO and MU-MIMO in the MIMO broadcast channel with CSI delay and channel quantization, and propose to switch between SU and MU-MIMO modes[1] based on the achievable rate of each technique with practical receiver assumptions. The main contributions in this chapter are as follows.

1. **SU-MIMO vs. MU-MIMO Analysis**. We investigate the impact of imperfect CSIT due to delay and channel quantization. We show that the SU mode is more robust to imperfect CSIT as it only suffers a constant rate loss, while MU-MIMO suffers more severely from residual inter-user interference. We characterize the residual interference due to delay and channel quantization, which shows these two effects are equivalent. Based on an independence approximation of the interference terms and the signal term, accurate closed-form approximations are derived for ergodic achievable rates for both SU and MU-MIMO modes.

2. **Mode Switching Algorithm.** An SU/MU mode switching algorithm is proposed based on the ergodic sum rate as a function of average SNR, normalized Doppler frequency, and the quantization codebook size. This transmission technique only requires a small number of users to feed back instantaneous channel information. The mode switching points can be calculated from the previously derived approximations for ergodic rates.

3. **Operating Regions.** *Operating regions* for SU and MU modes are determined, from which we can determine the active mode and find the condition that activates each mode. With a fixed delay and codebook size, if the MU mode is possible at all, there are two mode switching points, with the SU mode preferred at both low and high SNRs. The MU mode will only be activated when the normalized Doppler frequency is very small and the codebook size is large. From the numerical

[1]Note that "mode" in this Chapter refers to the single-user mode (SU-MIMO transmission) or multi-user mode (MU-MIMO transmission). This differs from use of the term in some related recent work (all for single user MIMO), for example switching between spatial multiplexing and diversity mode [149] or between different number of data streams per user [100, 117, 148].

Table 2.1: System Parameters

Symbol	Description
N_t	number of transmit antennas
U	number of mobile users
B	number of feedback bits
L	quantization codebook size, $L = 2^B$
P	average SNR
n	time index
T_s	the length of each symbol
f_d	the Doppler frequency

results, the minimum feedback bits per user to get the MU mode activated grow approximately linearly with the number of transmit antennas.

2.3 System Model

We consider a MIMO broadcast channel, where the transmitter (the BS) has N_t antennas and each mobile user has a single antenna. The system parameters are listed in Table 2.1. During each transmission period, which is less than the channel coherence time and the channel is assumed to be constant, the BS transmits to one (SU-MIMO mode) or multiple (MU-MIMO mode) users. For the MU-MIMO mode, we assume the number of active users is $U = N_t$, and the users are scheduled independently of their channel conditions, e.g. through round-robin scheduling, random user selection, or scheduling based on the queue length. The discrete-time complex baseband received signal at the u-th user at time n is given as[2]

$$y_u[n] = \mathbf{h}_u^*[n] \sum_{u'=1}^{U} \mathbf{f}_{u'}[n] x_{u'}[n] + z_u[n], \tag{2.1}$$

where $\mathbf{h}_u[n]$ is the $N_t \times 1$ channel vector from the transmitter to the u-th user, and $z_u[n]$ is the normalized complex Gaussian noise vector, i.e. $z_u[n] \sim \mathcal{CN}(0, 1)$. $x_u[n]$

[2]In this dissertation, we use uppercase boldface letters for matrices ($\mathbf{X}$) and lowercase boldface for vectors ($\mathbf{x}$). $\mathbb{E}[\cdot]$ is the expectation operator. The conjugate transpose of a matrix $\mathbf{X}$ (vector $\mathbf{x}$) is $\mathbf{X}^*$ ($\mathbf{x}^*$). Similarly, $\mathbf{X}^\dagger$ denotes the pseudo-inverse, $\tilde{\mathbf{x}}$ denotes the normalized vector of $\mathbf{x}$, i.e. $\tilde{\mathbf{x}} = \frac{\mathbf{x}}{\|\mathbf{x}\|}$, and $\hat{\mathbf{x}}$ denotes the quantized vector of $\tilde{\mathbf{x}}$.

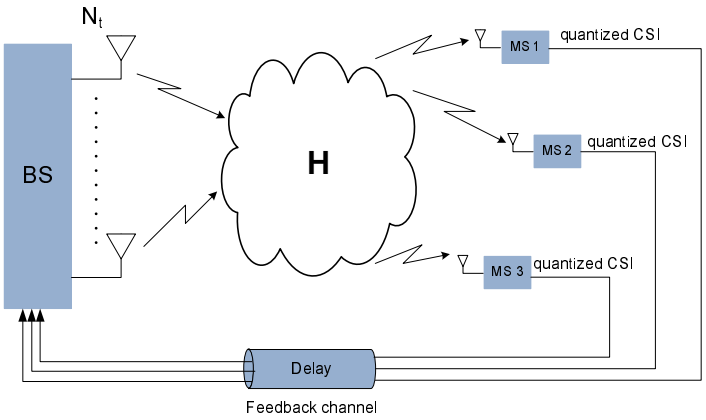


Figure 2.1: The system model for the MIMO broadcast channel with imperfect CSIT.

and $\mathbf{f}_u[n]$ are the transmit signal and the normalized $N_t \times 1$ precoding vector for the u-th user, respectively. The transmit power constraint is $\mathbb{E}\{\mathbf{x}^*[n]\mathbf{x}[n]\} = P$, where $\mathbf{x}[n] = [x_1^*, x_2^*, \cdots, x_U^*]^*$. As the noise is normalized, P is also the average transmit SNR.

To assist the analysis, we assume that the channel $\mathbf{h}_u[n]$ is well modeled as a spatially white Gaussian channel, with entries $h_{i,j}[n] \sim \mathcal{CN}(0,1)$, and the channels are i.i.d. over different users. Note that in the case of line of sight MIMO channel, fewer feedback bits are required compared to the Rayleigh channel [115].

The model for the MIMO broadcast channel with imperfect CSIT is shown in Figure 2.1. We consider two of the main sources of the CSIT imperfection–delay and quantization error[3], specified as follows.

2.3.1 CSI Delay Model

We consider a stationary ergodic Gauss-Markov block fading process [62, Sec. 16–1], where the channel stays constant for a symbol duration and changes from

[3]For a practical system, the feedback bits for each user is usually fixed, and there will inevitably be delay in the available CSI, both of which are difficult or even impossible to adjust. Other effects such as channel estimation error can be made small such as by increasing the transmit power or the number of pilot symbols.

symbol to symbol according to

$$\mathbf{h}[n] = \rho\mathbf{h}[n-1] + \mathbf{e}[n], \tag{2.2}$$

where $\mathbf{e}[n]$ is the channel error vector, with i.i.d. entries $e_i[n] \sim \mathcal{CN}(0, \epsilon_e^2)$, and it is uncorrelated with $\mathbf{h}[n-1]$. We assume the CSI delay is of one symbol. It is straightforward to extend the results to the scenario with a delay of multiple symbols. For the numerical analysis, the classical Clarke's isotropic scattering model will be used as an example, for which the correlation coefficient is $\rho = J_0(2\pi f_d T_s)$ with Doppler spread f_d [35], where $J_0(\cdot)$ is the zero-th order Bessel function of the first kind. The variance of the error vector is $\epsilon_e^2 = 1 - \rho^2$. Therefore, both ρ and ϵ_e are determined by the normalized Doppler frequency $f_d T_s$.

Remark 1. *This Gauss-Markov channel model is adopted as it is amenable to analysis and captures the essential characteristics of delay. The correlation coefficient ρ is determined by the amount of delay. Based on this model we are able to derive a closed-form expression for the achievable throughput, which, therefore, explicitly shows the impact of delay. The emphasis of this chapter is to show the impact of imperfect CSIT, and we have not exploited the stochastic property of the channel model, which can be used to further improve the performance, e.g. by exploiting the temporal correlation [69].*

The channel in (2.2) is widely-used to model the time-varying channel. For example, it is used to investigate the impact of feedback delay on the performance of closed-loop transmit diversity in [108] and the system capacity and bit error rate of point-to-point MIMO link in [107]. It simplifies the analysis, and the results can be easily extended to other scenarios with the channel model of the form

$$\mathbf{h}[n] = \mathbf{g}[n] + \mathbf{e}[n], \tag{2.3}$$

where $\mathbf{g}[n]$ is the available CSI at time n with an uncorrelated error vector $\mathbf{e}[n]$, $\mathbf{g}[n] \sim \mathcal{CN}(\mathbf{0}, (1-\epsilon_e^2)\mathbf{I})$, and $\mathbf{e}[n] \sim \mathcal{CN}(\mathbf{0}, \epsilon_e^2\mathbf{I})$. It can be used to consider the effect of other imperfect CSIT, such as estimation error and analog feedback. The difference is in $\mathbf{e}[n]$, which has different variance ϵ_e^2 for different scenarios. Some examples are given as follows.

Estimation Error If the receiver obtains the CSI through minimum mean-squared error (MMSE) estimation from τ_p pilot symbols, the error variance is $\epsilon_e^2 = \frac{1}{1+\tau_p\gamma_p}$, where γ_p is the SNR of the pilot symbol [21].

Analog Feedback For analog feedback, the error variance is $\epsilon_e^2 = \frac{1}{1+\tau_{ul}\gamma_{ul}}$, where τ_{ul} is the number of channel uses per channel coefficient and γ_{ul} is the SNR on the uplink feedback channel [22].

Analog Feedback with Prediction As shown in [89], for analog feedback with a d-step MMSE predictor and the Gauss-Markov model, the error variance is $\epsilon_e^2 = \rho^{2d}\epsilon_0 + (1-\rho^2)\sum_{l=0}^{d-1}\rho^{2l}$, where ρ is the same as in (2.2) and ϵ_0 is the Kalman filtering mean-square error.

Therefore, the results in this chapter can be easily extended to these systems. In the following parts, we focus on the effect of CSI delay.

2.3.2 Channel Quantization Model

We consider frequency-division duplexing (FDD) systems, where limited feedback techniques provide partial CSIT through a dedicated feedback channel from the receiver to the transmitter. The channel direction information for the precoder design is fed back using a quantization codebook known at both the transmitter and receiver.

The quantization is chosen from a codebook of unit norm vectors of size $L = 2^B$. We assume each user uses a different codebook to avoid the same quantization vector. The codebook for user u is $\mathcal{C}_u = \{\mathbf{c}_{u,1}, \mathbf{c}_{u,2}, \cdots, \mathbf{c}_{u,L}\}$. Each user quantizes its channel to the closest codeword, where closeness is measured by the inner product. Therefore, the index of channel for user u is

$$I_u = \arg\max_{1\le\ell\le L} |\tilde{\mathbf{h}}_u^* \mathbf{c}_{u,\ell}|. \tag{2.4}$$

Each user needs to feed back B bits to denote this index, and the transmitter has the quantized channel information $\hat{\mathbf{h}}_u = \mathbf{c}_{u,I_u}$. As the optimal vector quantizer for this

problem is not known in general, random vector quantization (RVQ) [120] is used, where each quantization vector is independently chosen from the isotropic distribution on the N_t-dimensional unit sphere. It has been shown in [75] that RVQ can facilitate the analysis and provide performance close to the optimal quantization. In this chapter, we analyze the achievable rate averaged over both RVQ-based random codebooks and fading distributions.

An important metric for the limited feedback system is the squared angular distortion, defined as $\sin^2(\theta_u) = 1 - |\tilde{\mathbf{h}}_u^* \hat{\mathbf{h}}_u|^2$, where $\theta_u = \angle\left(\tilde{\mathbf{h}}_u, \hat{\mathbf{h}}_u\right)$. With RVQ, it was shown in [13, 75] that the expectation in i.i.d. Rayleigh fading is given by

$$\mathbb{E}_\theta\left[\sin^2(\theta_u)\right] = 2^B \cdot \beta\left(2^B, \frac{N_t}{N_t - 1}\right), \tag{2.5}$$

where $\beta(\cdot)$ is the beta function [58]. It can be tightly bounded as [75]

$$\frac{N_t - 1}{N_t} 2^{-\frac{B}{N_t - 1}} \leq \mathbb{E}\left[\sin^2(\theta_u)\right] \leq 2^{-\frac{B}{N_t - 1}}. \tag{2.6}$$

2.4 Transmission Techniques

In this section, we describe the transmission techniques for both SU and MU-MIMO systems with perfect CSIT, which will be used in the subsequent sections for imperfect CSIT systems. By doing this, we focus on the impacts of imperfect CSIT on the conventional transmission techniques. Throughout this chapter, we use the achievable ergodic rate as the performance metric for both SU and MU-MIMO systems. The base station transmits to a single user ($U = 1$) for the SU-MIMO system and to N_t users ($U = N_t$) for the MU-MIMO system. The SU/MU mode switching algorithm is also described.

2.4.1 SU-MIMO System

With perfect CSIT, it is optimal for the SU-MIMO system to transmit along the channel direction [136], i.e. selecting the beamforming (BF) vector as $\mathbf{f}[n] = \tilde{\mathbf{h}}[n]$, denoted as *eigen-beamforming*. The ergodic capacity of this system is the same as

that of a maximal ratio combining diversity system, given by [10]

$$R_{BF}(P) = \mathbb{E}_{\mathbf{h}}\left[\log_2\left(1 + P\|\mathbf{h}[n]\|^2\right)\right]$$
$$= \log_2(e)e^{1/P}\sum_{k=0}^{N_t-1}\frac{\Gamma(-k, 1/P)}{P^k}, \tag{2.7}$$

where $\Gamma(\cdot,\cdot)$ is the complementary incomplete gamma function defined as $\Gamma(\alpha, x) = \int_x^\infty t^{\alpha-1}e^{-t}\mathrm{d}t$.

2.4.2 MU-MIMO System

For the multi-antenna broadcast channel, although dirty-paper coding (DPC) [36] is optimal [151], it is difficult to implement in practice. As in [21, 75], ZF precoding is used in this chapter, which is a linear precoding technique that precancels inter-user interference at the transmitter. There are several reasons for us to use this simple transmission technique. Firstly, due to its simple structure, it is possible to derive closed-form results, which can provide helpful insights. Second, the ZF precoding is able to provide full spatial multiplexing gain and only has a power offset compared to the optimal DPC system [74]. In addition, it was shown in [74] that the ZF precoding is optimal among the set of all linear precoders at asymptotically high SNR. In Section 2.6, we will show that our results for the ZF system also apply for the regularized ZF precoding (aka MMSE precoding) [113], which provides a higher throughput than the ZF precoding at low to moderate SNRs.

With precoding vectors $\mathbf{f}_u[n], u = 1, 2, \cdots, U$, assuming equal power allocation[4], the received SINR for the u-th user is given as

$$\gamma_{ZF,u} = \frac{\frac{P}{U}|\mathbf{h}_u^*[n]\mathbf{f}_u[n]|^2}{1 + \frac{P}{U}\sum_{u'\neq u}|\mathbf{h}_u^*[n]\mathbf{f}_{u'}[n]|^2}.$$

This is true for a general linear precoding MU-MIMO system. With perfect CSIT, this quantity can be calculated at the transmitter, while with imperfect CSIT, it can be estimated at the receiver and fed back to the transmitter given knowledge of $\mathbf{f}_u[n]$.

[4]At high SNR, this performs closely to the system employing optimal water-filling, as power allocation mainly benefits at low SNR.

Denote $\tilde{\mathbf{H}}[n] = [\tilde{\mathbf{h}}_1[n], \tilde{\mathbf{h}}_2[n], \cdots, \tilde{\mathbf{h}}_U[n]]^*$. With perfect CSIT, the ZF precoding vectors are determined from the pseudo-inverse of $\tilde{\mathbf{H}}[n]$, as $\mathbf{F}[n] = \tilde{\mathbf{H}}^\dagger[n] = \tilde{\mathbf{H}}^*[n](\tilde{\mathbf{H}}[n]\tilde{\mathbf{H}}^*[n])^{-1}$. The precoding vector for the u-th user is obtained by normalizing the u-th column of $\mathbf{F}[n]$. Therefore, $\mathbf{h}_u^*[n]\mathbf{f}_{u'}[n] = 0$, $\forall u \neq u'$, i.e. there is no inter-user interference. The received SINR for the u-th user becomes

$$\gamma_{ZF,u} = \frac{P}{U}|\mathbf{h}_u^*[n]\mathbf{f}_u[n]|^2. \tag{2.8}$$

As $\mathbf{f}_u[n]$ is independent of $\mathbf{h}_u[n]$, and $\|\mathbf{f}_u[n]\|^2 = 1$, the effective channel for the u-th user is a single-input single-output (SISO) Rayleigh fading channel. Therefore, the achievable sum rate for the ZF system is given by

$$R_{ZF}(P) = \sum_{u=1}^{U} \mathbb{E}_\gamma \left[\log_2(1 + \gamma_{ZF,u})\right]. \tag{2.9}$$

Each term on the right hand side of (2.9) is the ergodic capacity of a SISO system in Rayleigh fading, given in [10] as

$$\begin{aligned} R_{ZF,u} &= \mathbb{E}_\gamma \left[\log_2(1 + \gamma_{ZF,u})\right] \\ &= \log_2(e)e^{U/P}E_1(U/P), \end{aligned} \tag{2.10}$$

where $E_1(\cdot)$ is the exponential-integral function of the first order, $E_1(x) = \int_1^\infty \frac{e^{-xt}}{t}\mathrm{d}t$.

2.4.3 SU/MU Mode Switching

Imperfect CSIT will degrade the performance of the MIMO communication. In this case, it is unclear whether and when the MU-MIMO system can actually provide a throughput gain over the SU-MIMO system. Based on the analysis of the achievable ergodic rates in this chapter, we propose to switch between SU and MU modes and select the one with the higher achievable rate.

The channel correlation coefficient ρ, which captures the CSI delay effect, usually varies slowly. The quantization codebook size is normally fixed for a given system. Therefore, it is reasonable to assume that the transmitter has knowledge of both delay and channel quantization, and can estimate the achievable ergodic rates

of both SU and MU-MIMO modes. Then it can determine the active mode and select one (SU mode) or N_t (MU mode) users to serve. This is a low-complexity transmission strategy, and can be combined with random user selection, round-robin scheduling, or scheduling based on queue length rather than channel status. It only requires the selected users to feed back instantaneous channel information. Therefore, it is suitable for a system that has a constraint on the total feedback bits and only allows a small number of users to send feedback, or a system with a strict delay constraint that cannot employ opportunistic scheduling based on instantaneous channel information.

To determine the transmission rate, the transmitter sends pilot symbols, from which the active users estimate the received SINRs and feed back them to the transmitter. In this research, we assume the transmitter knows perfectly the actual received SINR at each active user, so there will be no outage in the transmission.

2.5 SU vs. MU with Delayed and Quantized CSIT

In this section, we investigate the achievable ergodic rates for both SU and MU-MIMO modes. We first analyze the average received SNR for the BF system and the average residual interference for the ZF system, which provide insights on the impact of imperfect CSIT. To select the active mode, accurate closed-form approximations for both SU and MU modes are then derived.

2.5.1 SU Mode–Eigen-Beamforming

First, if there is no delay and only channel quantization, the BF vector is based on the quantized feedback, $\mathbf{f}^{(Q)}[n] = \hat{\mathbf{h}}[n]$[5]. The average received SNR is

$$\begin{aligned}\overline{\mathrm{SNR}}_{BF}^{(Q)} &= \mathbb{E}_{\mathbf{h},\mathrm{e}}[P|\mathbf{h}^*[n]\hat{\mathbf{h}}[n]|^2] \\ &= \mathbb{E}_{\mathbf{h},\mathrm{e}}[P\|\mathbf{h}[n]\|^2|\tilde{\mathbf{h}}^*[n]\hat{\mathbf{h}}[n]|^2] \\ &\overset{(a)}{\leq} PN_t\left(1-\frac{N_t-1}{N_t}2^{-\frac{B}{N_t-1}}\right), \end{aligned} \tag{2.11}$$

[5]Letters Q and D denote parameters of the system with quantization (limited feedback) and delay, respectively.

where (a) follows by the independence between $\|\mathbf{h}[n]\|^2$ and $|\tilde{\mathbf{h}}^*[n]\hat{\mathbf{h}}[n]|^2$, together with the result in (2.6).

With both delay and channel quantization, the BF vector is based on the quantized channel direction with delay, i.e. $\mathbf{f}^{(QD)}[n] = \hat{\mathbf{h}}[n-1]$. The instantaneous received SNR for the BF system

$$\mathrm{SNR}_{BF}^{(QD)} = P\left|\mathbf{h}^*[n]\mathbf{f}^{(QD)}[n]\right|^2. \tag{2.12}$$

Based on (2.11), we get the following theorem on the average received SNR for the SU mode.

Theorem 1. *The average received SNR for a BF system with channel quantization and CSI delay is*

$$\overline{SNR}_{BF}^{(QD)} \leq PN_t\left(\rho^2\Delta_{BF}^{(Q)} + \Delta_{BF}^{(D)}\right), \tag{2.13}$$

where $\Delta_{BF}^{(Q)}$ and $\Delta_{BF}^{(D)}$ show the impact of channel quantization and feedback delay, respectively, given by

$$\Delta_{BF}^{(Q)} = 1 - \frac{N_t - 1}{N_t}2^{-\frac{B}{N_t-1}}, \quad \Delta_{BF}^{(D)} = \frac{\epsilon_e^2}{N_t}.$$

Proof. See Appendix 2.8.2. □

From Jensen's inequality, an upper bound of the achievable rate for the BF system with both quantization and delay is given by

$$\begin{aligned} R_{BF}^{(QD)} &= \mathbb{E}_{\mathbf{h},\mathfrak{e}}\left[\log_2\left(1+\mathrm{SNR}_{BF}^{(QD)}\right)\right] \\ &\leq \log_2\left[1+\overline{\mathrm{SNR}}_{BF}^{(QD)}\right] \\ &\leq \log_2\left[1+PN_t\left(\rho^2\Delta_{BF}^{(Q)}+\Delta_{BF}^{(D)}\right)\right]. \end{aligned} \tag{2.14}$$

Remark 2. *Note that $\rho^2 = 1-\epsilon_e^2$, so the average SNR decreases with ϵ_e^2. With a fixed B and fixed delay, the SNR degradation is a constant factor independent of P. At high SNR, the imperfect CSIT introduces a constant rate loss $\log_2\left(\rho^2\Delta_{BF}^{(Q)}+\Delta_{BF}^{(D)}\right)$.*

The upper bound provided by Jensen's inequality is not tight. To get a better approximation for the achievable rate, we first make the following approximation on the instantaneous received SNR

$$\begin{aligned}\mathrm{SNR}_{BF}^{(QD)} &= P|\mathbf{h}^*[n]\hat{\mathbf{h}}[n-1]|^2 \\ &= P|(\rho\mathbf{h}[n-1]+\mathbf{e}[n])^*\hat{\mathbf{h}}[n-1]|^2 \\ &\approx P\rho^2|\mathbf{h}^*[n-1]\hat{\mathbf{h}}[n-1]|^2, \end{aligned} \tag{2.15}$$

i.e. we remove the term with $\mathbf{e}[n]$ as it is normally insignificant compared to $\rho\mathbf{h}[n-1]$. This will be verified later by simulation. In this way, the system is approximated as the one with limited feedback and with equivalent SNR $\rho^2 P$.

From [13], the achievable rate of the limited feedback BF system is given by

$$\begin{aligned}R_{BF}^{(Q)}(P) = \log_2(e)\Bigg(&e^{1/P}\sum_{k=0}^{N_t-1}E_{k+1}\left(\frac{1}{P}\right) \\ &-\int_0^1\left(1-(1-x)^{N_t-1}\right)^{2^B}\frac{N_t}{x}e^{1/Px}E_{N_t+1}\left(\frac{1}{Px}\right)\mathrm{d}x\Bigg),\end{aligned} \tag{2.16}$$

where $E_n(x)=\int_1^\infty e^{-xt}x^{-n}\mathrm{d}t$ is the n-th order exponential integral. So $R_{BF}^{(QD)}$ can be approximated as

$$R_{BF}^{(QD)}(P) \approx R_{BF}^{(Q)}(\rho^2 P). \tag{2.17}$$

As a special case, considering a system with delay only, e.g. the time-division duplexing (TDD) system which can estimate the CSI from the uplink with channel reciprocity but with propagation and processing delay, the BF vector is based on the delayed channel direction, i.e. $\mathbf{f}^{(D)}[n]=\tilde{\mathbf{h}}[n-1]$. We provide a good approximation for the achievable rate for such a system as follows.

The instantaneous received SNR is given as

$$\begin{aligned}\mathrm{SNR}_{BF}^{(D)} &= P|\mathbf{h}^*[n]\mathbf{f}^{(D)}[n]|^2 \\ &= P|(\rho\mathbf{h}[n-1]+\mathbf{e}[n])^*\tilde{\mathbf{h}}[n-1]|^2 \\ &\overset{(a)}{\approx} P\rho^2\|\mathbf{h}[n-1]\|^2 + P|\mathbf{e}^*[n]\tilde{\mathbf{h}}[n-1]|^2. \end{aligned} \tag{2.18}$$

In step (a) we eliminate the cross terms since $\mathbf{e}[n]$ is normally small, e.g. its various is $\epsilon_e^2 = 0.027$ with carrier frequency at 2 GHz, mobility of 20 km/hr and delay of 1 msec. As $\mathbf{e}[n]$ is independent of $\tilde{\mathbf{h}}[n-1]$, $\mathbf{e}[n] \sim \mathcal{CN}(\mathbf{0}, \epsilon_e^2\mathbf{I})$ and $\|\tilde{\mathbf{h}}[n-1]\|^2 = 1$, we have $|\mathbf{e}^*[n]\tilde{\mathbf{h}}[n-1]|^2 \sim \chi_2^2$, where χ_M^2 denotes chi-square distribution with M degrees of freedom. In addition, $\|\mathbf{h}[n-1]\|^2 \sim \chi_{2N_t}^2$, and it is independent of $|\mathbf{e}^*[n]\tilde{\mathbf{h}}[n-1]|^2$. Then the following theorem can be derived.

Theorem 2. *The achievable ergodic rate of the BF system with delay can be approximated as*

$$R_{BF}^{(D)} \approx \log_2(e) {a_0}^{N_t} e^{1/\eta_2} E_1\left(\frac{1}{\eta_2}\right) - \log_2(e)(1-a_0)\sum_{i=0}^{N_t-1}\sum_{l=0}^{i}\frac{a_0^{N_t-1-i}}{(i-l)!}\eta_1^{-(i-l)} I_1(1/\eta_1, 1, i-l), \tag{2.19}$$

where $\eta_1 = P\rho^2$, $\eta_2 = P\epsilon_e^2$, $a_0 = \frac{\eta_2}{\eta_2-\eta_1}$, *and* $I_1(\cdot,\cdot,\cdot)$ *is given in* (2.35) *in Appendix 2.8.1.*

Proof. See Appendix 2.8.3. □

2.5.2 MU Mode–Zero-Forcing

If there is no delay but only channel quantization, the precoding vectors for the ZF system are designed based on $\hat{\mathbf{h}}_1[n], \hat{\mathbf{h}}_2[n], \cdots, \hat{\mathbf{h}}_U[n]$ to achieve $\hat{\mathbf{h}}_u^*[n]\mathbf{f}_{u'}^{(Q)}[n] = 0$, $\forall u \neq u'$. With random vector quantization, it is shown in [75] that the average noise plus interference for each user is

$$\Delta_{ZF,u}^{(Q)} = \mathbb{E}_{\mathbf{h},\mathcal{C}}\left[1 + \frac{P}{U}\sum_{u'\neq u}|\mathbf{h}_u^*[n]\mathbf{f}_{u'}^{(Q)}[n]|^2\right] = 1 + 2^{-\frac{B}{N_t-1}}P. \tag{2.20}$$

With both channel quantization and CSI delay, precoding vectors are designed based on $\hat{\mathbf{h}}_1[n-1], \hat{\mathbf{h}}_2[n-1], \cdots, \hat{\mathbf{h}}_U[n-1]$ and achieve $\hat{\mathbf{h}}_u^*[n-1]\mathbf{f}_{u'}^{(QD)}[n] = 0$, $\forall u \neq u'$. The received SINR for the u-th user is given as

$$\gamma_{ZF,u}^{(QD)} = \frac{\frac{P}{U}|\mathbf{h}_u^*[n]\mathbf{f}_u^{(QD)}[n]|^2}{1 + \frac{P}{U}\sum_{u'\neq u}|\mathbf{h}_u^*[n]\mathbf{f}_{u'}^{(QD)}[n]|^2}. \tag{2.21}$$

As $\mathbf{f}_u^{(QD)}[n]$ is in the nullspace of $\hat{\mathbf{h}}_{u'}[n-1]$ $\forall u' \neq u$, it is isotropically distributed in $\mathbb{C}^{N_t}$ and independent of $\tilde{\mathbf{h}}_u[n-1]$ as well as $\tilde{\mathbf{h}}_u[n]$, so $|\mathbf{h}_u^*[n]\mathbf{f}_u^{(QD)}[n]|^2 \sim \chi_2^2$. The average noise plus interference is given in the following theorem.

Theorem 3. *The average noise plus interference for the u-th user of the ZF system with both channel quantization and CSI delay is*

$$\Delta_{ZF,u}^{(QD)} = 1 + (U-1)\frac{P}{U}\left(\rho_u^2 \Delta_{ZF,u}^{(Q)} + \Delta_{ZF,u}^{(D)}\right), \tag{2.22}$$

where $\Delta_{ZF,u}^{(Q)}$ *and* $\Delta_{ZF,u}^{(D)}$ *are the degradations brought by channel quantization and feedback delay, respectively, given by*

$$\Delta_{ZF,u}^{(Q)} = \frac{U}{U-1} 2^{-\frac{B}{N_t-1}}, \quad \Delta_{ZF,u}^{(D)} = \epsilon_{e,u}^2.$$

Proof. See Appendix 2.8.4. □

Remark 3. *From* Theorem 3 *we see that the average residual interference for a given user consists of three parts:*

(i) The number of interferers, $U-1$. *The more users the system supports, the higher the mutual interference.*

(ii) The transmit power of the other active users, $\frac{P}{U}$. *As the transmit power increases, the system becomes interference-limited.*

(iii) The CSIT accuracy for this user, *which is reflected from* $\rho_u^2 \Delta_{ZF,u}^{(Q)} + \Delta_{ZF,u}^{(D)}$. *The user with a larger delay or a smaller codebook size suffers a higher residual interference.*

From this remark, the interference term, $\frac{P}{U}(U-1)\epsilon_{e,u}^2$, equivalently comes from $U-1$ *virtual interfering users*, each with equivalent SNR as $\frac{P}{U}\left(\rho_u^2 \Delta_{ZF,u}^{(Q)} + \Delta_{ZF,u}^{(D)}\right)$. With a high P and a fixed $\epsilon_{e,u}$ or B, the system is interference-limited and cannot achieve the full spatial multiplexing gain. Therefore, to keep a constant rate loss, i.e. to sustain the spatial multiplexing gain, the channel error due to both quantization and delay needs to be reduced as SNR increases. Similar to the result for the limited

feedback system in [75], for the ZF system with both delay and channel quantization, we can get the following corollary for the condition to achieve the full spatial multiplexing gain.

Corollary 1. *To keep a constant rate loss of* $\log_2 \delta_0$ *bps/Hz for each user, the codebook size and CSI delay need to satisfy the following condition*

$$\rho_u^2 \Delta_{ZF,u}^{(Q)} + \Delta_{ZF,u}^{(D)} = \frac{U}{U-1} \cdot \frac{\delta_0 - 1}{P}. \tag{2.23}$$

Proof. As shown in [21, 75], the rate loss for each user due to imperfect CSIT is upper bounded by $\Delta R_u \leq \log_2 \Delta_{ZF,u}^{(QD)}$. The corollary follows from solving $\log_2 \Delta_{ZF,u}^{(QD)} = \log_2 \delta_0$. □

Equivalently, this means that for a given ρ^2, the feedback bits per user needs to scale as

$$B = (N_t - 1) \log_2 \left(\frac{\delta_0 - 1}{\rho_u^2 P} - \frac{U-1}{U} \cdot \left(\frac{1}{\rho_u^2} - 1 \right) \right)^{-1}. \tag{2.24}$$

As $\rho_u^2 \to 1$, i.e. there is no CSI delay, the condition becomes $B = (N_t - 1) \log_2 \frac{P}{\delta_0 - 1}$, which agrees with the result in [75] with limited feedback only.

For the ZF system with imperfect CSI, the genie-aided upper bound for the ergodic achievable rate[6] is given by [21]

$$R_{ZF}^{(QD)} \leq \sum_{u=1}^{U} \mathbb{E}_\gamma \left[\log_2 \left(1 + \gamma_{ZF,u}^{(QD)} \right) \right] = R_{ZF,ub}^{(QD)}. \tag{2.25}$$

We assume the mobile users can perfectly estimate the noise and interference and feed back it to the transmitter, so the upper bound is chosen as the performance metric, i.e. $R_{ZF}^{(QD)} = R_{ZF,ub}^{(QD)}$, as in [43, 75, 156].

The following lower bound based on the rate loss analysis is used in [21, 75]

$$R_{ZF}^{(QD)} \geq R_{ZF} - \sum_{u=1}^{U} \log_2 \Delta_{ZF,u}^{(QD)}, \tag{2.26}$$

[6]This upper bound is achievable only when a genie provides users with perfect knowledge of all interference and the transmitter knows perfectly the received SINR at each user.

where R_{ZF} is the achievable rate with perfect CSIT, given in (2.9). However, this lower bound is very loose. In the following, we will derive a more accurate approximation for the achievable rate for the ZF system.

To get a good approximation for the achievable rate for the ZF system, we first approximate the instantaneous SINR as

$$\begin{aligned}\gamma_{ZF,u}^{(QD)} &= \frac{\frac{P}{U}|\mathbf{h}_u^*[n]\mathbf{f}_u^{(QD)}[n]|^2}{1+\frac{P}{U}\sum_{u'\neq u}|(\rho_u\mathbf{h}_u[n-1]+\mathbf{e}_u[n])^*\mathbf{f}_{u'}^{(QD)}[n]|^2} \\ &\approx \frac{\frac{P}{U}|\mathbf{h}_u^*[n]\mathbf{f}_u^{(QD)}[n]|^2}{1+\frac{P}{U}\left(\sum_{u'\neq u}\rho_u^2|\mathbf{h}_u^*[n-1]\mathbf{f}_{u'}^{(QD)}[n]|^2+\sum_{u'\neq u}|\mathbf{e}_u^*[n]\mathbf{f}_{u'}^{(QD)}[n]|^2\right)}, \end{aligned} \tag{2.27}$$

i.e. eliminating the interference terms which have both $\mathbf{h}_u[n-1]$ and $\mathbf{e}_u[n]$ as $\mathbf{e}_u[n]$ is normally very small, so we get two separate interference sums due to delay and quantization, respectively.

For the interference term due to delay, $|\mathbf{e}_u^*[n]\mathbf{f}_{u'}^{(QD)}[n]|^2 \sim \chi_2^2$, as $\mathbf{e}[n]$ is independent of $\mathbf{f}_{u'}^{(QD)}[n]$ and $\|\mathbf{f}_{u'}^{(QD)}[n]\|^2 = 1$. For the interference term due to quantization, it was shown in [75] that $|\tilde{\mathbf{h}}_u^*[n-1]\mathbf{f}_{u'}^{(QD)}[n]|^2$ is equivalent to the product of the quantization error $\sin^2\theta_u$ and an independent $\beta(1, N_t-2)$ random variable. Therefore, we have

$$|\mathbf{h}_u^*[n-1]\mathbf{f}_{u'}^{(QD)}[n] = \|\mathbf{h}_u[n-1]\|^2(\sin^2\theta_u)\cdot\beta(1,N_t-2). \tag{2.28}$$

In [156], with a quantization cell approximation[7] [103, 172], it was shown that $\|\mathbf{h}_u[n-1]\|^2(\sin^2\theta_u)$ has a Gamma distribution with parameters (N_t-1,δ), where $\delta = 2^{-\frac{B}{N_t-1}}$. As shown in [156] the analysis based on the quantization cell approximation is close to the performance of random vector quantization, so we use this approach to derive the achievable rate.

The following lemma gives the distribution of the interference term due to quantization.

[7]The quantization cell approximation is based on the ideal assumption that each quantization cell is a Voronoi region on a spherical cap with the surface area 2^{-B} of the total area of the unit sphere for a B bits codebook. The detail can be found in [103, 156, 172].

Lemma 1. *Based on the quantization cell approximation, the interference term due to quantization in* (2.27), $|\mathbf{h}_u[n-1]\mathbf{f}_{u'}^{(QD)}[n]|^2$, *is an exponential random variable with mean* δ, *i.e. its probability distribution function (pdf) is*

$$p(x) = \frac{1}{\delta} e^{-x/\delta}, x \geq 0. \tag{2.29}$$

Proof. See Appendix 2.8.5. □

Remark 4. *From this lemma, we see that the residual interference terms due to both delay and quantization are exponential random variables, which means the delay and quantization error have equivalent effects, only with different means. By comparing the means of these two terms, i.e. comparing* ϵ_e^2 *and* $2^{-\frac{B}{N_t-1}}$, *we can find the dominant one. In addition, with this result, we can approximate the achievable rate of the ZF limited feedback system, which will be provided later in this section.*

Based on the distribution of the interference terms, the approximation for the achievable rate for the MU mode is given in the following theorem.

Theorem 4. *The ergodic achievable rate for the u-th user in the MU mode with both delay and channel quantization can be approximated as*

$$R_{ZF,u}^{(QD)} \approx \log_2(e) \sum_{i=0}^{M-1} \sum_{j=1}^{2} \left[a_i^{(j)} i! \cdot I_3 \left(\frac{1}{\alpha}, \frac{1}{\delta_j}, i+1 \right) \right], \tag{2.30}$$

where $\alpha = \frac{P}{U}$, $\delta_1 = \rho_u^2 \delta$, $\delta_2 = \epsilon_{e,u}^2$, $M = N_t - 1$, $a_i^{(1)}$ *and* $a_i^{(2)}$ *are given in* (2.43) *and* (2.44), *and* $I_3(\cdot,\cdot,\cdot)$ *is given in* (2.37) *in Appendix 2.8.1.*

Proof. See Appendix 2.8.6. □

The ergodic sum throughput is

$$R_{ZF}^{(QD)} = \sum_{u=1}^{U} R_{ZF,u}^{(QD)}. \tag{2.31}$$

As a special case, for a ZF system with delay only, we can get the following approximation for the ergodic achievable rate.

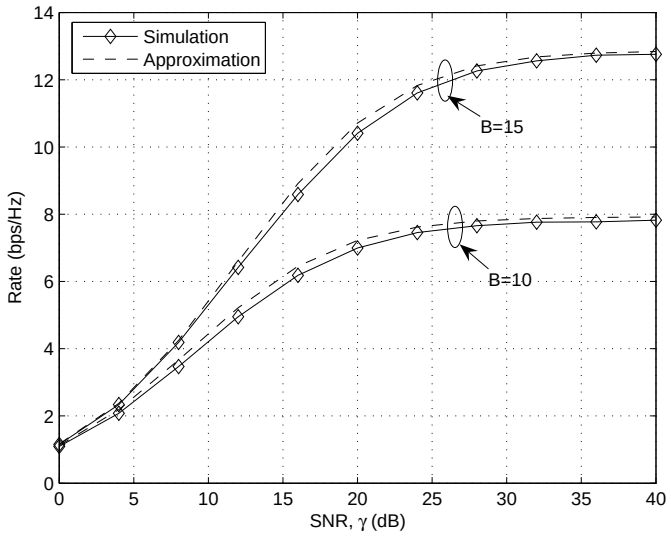


Figure 2.2: Approximated and simulated ergodic rates for the ZF precoding system with limited feedback, $N_t = U = 4$.

Corollary 2. *The ergodic achievable rate for the u-th user in the ZF system with delay is approximated as*

$$R_{ZF,u}^{(D)} \approx \log_2(e)\epsilon_{e,u}^{2(M-1)} \cdot I_3\left(\frac{1}{\alpha}, \frac{1}{\epsilon_{e,u}^2}, M-1\right), \tag{2.32}$$

where $\alpha = \frac{P}{U}$, $M = N_t - 1$, *and* $I_3(\cdot,\cdot,\cdot)$ *is given in* (2.37) *in Appendix 2.8.1.*

Proof. Following the same steps in Appendix 2.8.6 with $\delta_1 = 0$. □

Remark 5. *As shown in* Lemma 1*, the effects of delay and channel quantization are equivalent, so the approximation in* (2.32) *also applies for the limited feedback system. This is verified by simulation in Figure 2.2, which shows that this approximation is very accurate and can be used to analyze the limited feedback system.*

2.5.3 Mode Switching

We first verify the approximation (2.30) in Figure 2.3, which compares the approximation with simulation results and the lower bound (2.26), with $B = 10$ bits, $v = 20$ km/hr, $f_c = 2$ GHz, and $T_s = 1$ msec. We see that the lower bound is

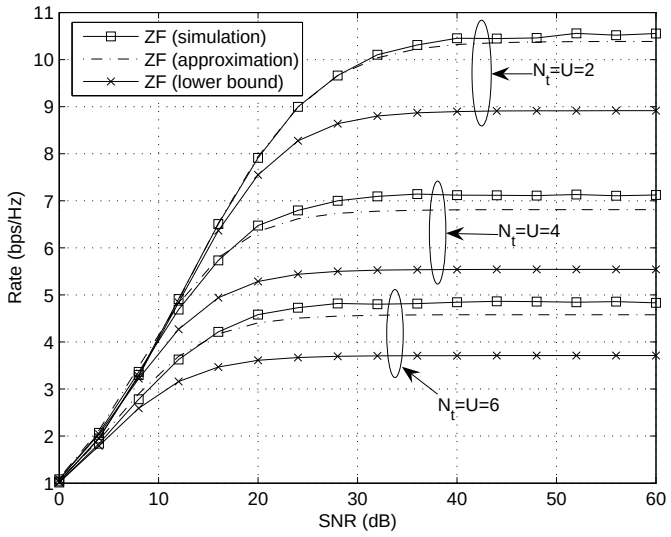


Figure 2.3: Comparison of approximation in (2.30), the lower bound in (2.26), and the simulation results for the ZF system with both delay and channel quantization. $B = 10$ bits, $f_c = 2$ GHz, $v = 20$ km/hr, and $T_s = 1$ msec.

very loose, while the approximation is accurate especially for $N_t = 2$. In fact, the approximation turns out to be a lower bound. Note that due to the imperfect CSIT, the sum rate reduces with N_t.

In Figure 2.4, we compare the BF and ZF systems, with $B = 18$ bits, $f_c = 2$ GHz, $v = 10$ km/hr, and $T_s = 1$ msec. We see that the approximation for the BF system almost matches the simulation exactly. The approximation for the ZF system is accurate at low to medium SNRs, and becomes a lower bound at high SNR, which is approximately 0.7 bps/Hz in total, or 0.175 bps/Hz per user, lower than the simulation. The throughput of the ZF system is limited by the residual inter-user interference at high SNR, where it is lower than the BF system. This motivates to switch between the SU and MU-MIMO modes. The approximations (2.17) and (2.30) will be used to calculate the mode switching points. There may be two switching points for the system with delay, as the SU mode will be selected at both low and high SNR. These two points can be calculated by providing different initial values to the nonlinear equation solver, such as *fsolve* in MATLAB.

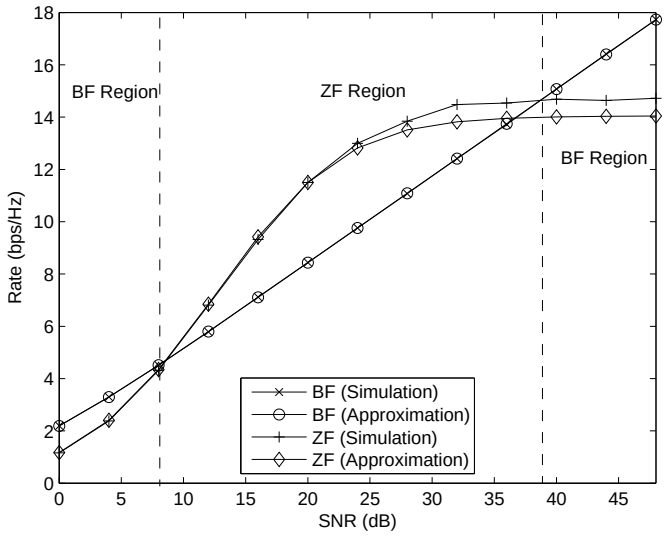


Figure 2.4: Mode switching between BF and ZF modes with both CSI delay and channel quantization, $B = 18$ bits, $N_t = 4$, $f_c = 2$ GHz, $T_s = 1$ msec, $v = 10$ km/hr.

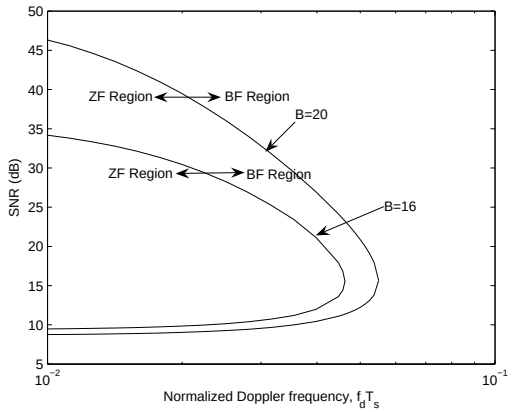
2.6 Numerical Results

In this section, numerical results are presented. First, the operating regions for different modes are plotted, which show the impact of different parameters, including the normalized Doppler frequency, the codebook size, and the number of transmit antennas. Then the extension of our results for ZF precoding to MMSE precoding is demonstrated.

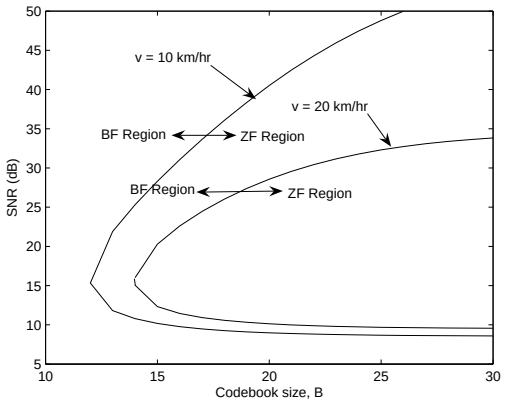
2.6.1 Operating Regions

As shown in Section 2.5.3, finding mode switching points requires solving a nonlinear equation, which does not have a closed-form solution and gives little insight. However, it is easy to evaluate numerically for different parameters, from which insights can be drawn. In this section, with the calculated mode switching points for different parameters, we plot the operating regions for both SU and MU modes. The active mode for the given parameter and the condition to activate each mode can be found from such plots.

In Figure 2.5, the operating regions for both SU and MU modes are plotted,



(a) Different f_dT_s.



(b) Different B, $f_c = 2$ GHz, $T_s = 1$ msec.

Figure 2.5: Operating regions for BF and ZF with both CSI delay and quantization, $N_t = 4$.

for different normalized Doppler frequencies and different number of feedback bits in Figure 2.5(a) and Figure 2.5(b), respectively, and with $U = N_t = 4$. There are analogies between the two plots. Some key observations are as follows:

1. For the delay plot Figure 2.5(a), comparing the two curves for $B = 16$ bits and $B = 20$ bits, we see that the smaller the codebook size, the smaller the operating region for the ZF mode. For the ZF mode to be active, f_dT_s needs to be small, specifically we need $f_dT_s < 0.055$ and $f_dT_s < 0.046$ for $B = 20$ bits and $B = 16$ bits, respectively. These conditions are not easily satisfied in practical systems. For example, with carrier frequency $f_c = 2$ GHz, mobility $v = 20$ km/hr, the Doppler frequency is 37 Hz, and then to satisfy $f_dT_s < 0.055$ the delay should be less than 1.5 msec.

2. For the codebook size plot Figure 2.5(b), comparing the two curves with $v = 10$ km/hr and $v = 20$ km/hr, as f_dT_s increases (v increases), the ZF operating region shrinks. For the ZF mode to be active, we should have $B \geq 12$ bits and $B \geq 14$ bits for $v = 10$ km/hr and $v = 20$ km/hr, respectively, which means a large codebook size. Note that for BF we only need a small codebook size to get the near-optimal performance [98].

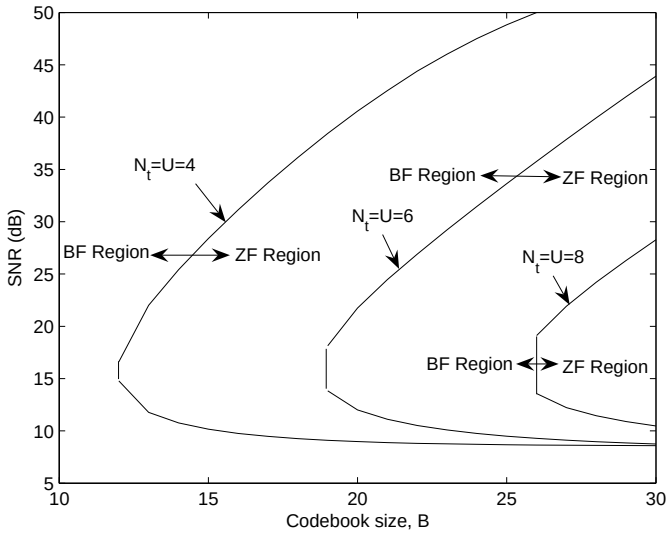


Figure 2.6: Operating regions for BF and ZF with different N_t, $f_c = 2$ GHz, $v = 10$ km/hr, $T_s = 1$ msec.

3. For a given $f_d T_s$ and B, the SU mode will be active at both low and high SNRs, which is due to its array gain and the robustness to imperfect CSIT, respectively.

The operating regions for different N_t are shown in Figure 2.6. We see that as N_t increases, the operating region for the MU mode shrinks. Specifically, we need $B > 12$ bits for $N_t = 4$, $B > 19$ bits for $N_t = 6$, and $B > 26$ bits for $N_t = 8$ to get the MU mode activated. Note that the minimum required feedback bits per user for the MU mode grow approximately linearly with N_t.

2.6.2 ZF vs. MMSE Precoding

It is shown in [113] that the regularized ZF precoding, denoted as *MMSE precoding*, can significantly increase the throughput at low SNR. In this section, we show that our results on mode switching with ZF precoding can also be applied to MMSE precoding.

Denote $\hat{\mathbf{H}}[n] = \left[\hat{\mathbf{h}}_1[n], \hat{\mathbf{h}}_2[n], \cdots, \hat{\mathbf{h}}_U[n]\right]^*$. Then the MMSE precoding vectors

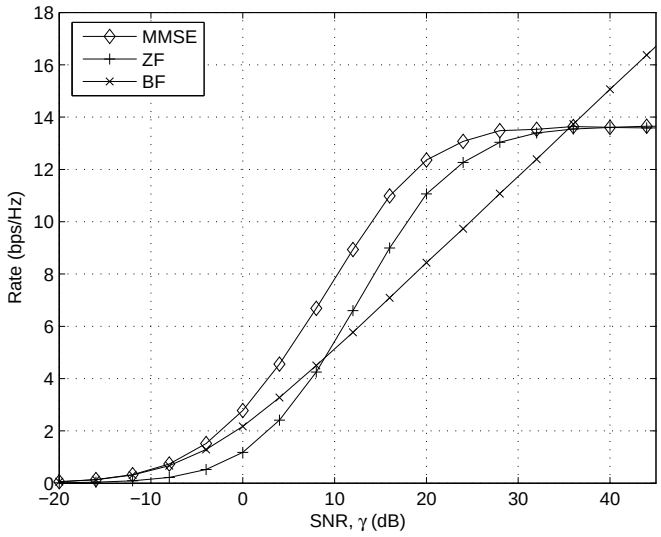


Figure 2.7: Simulation results for BF, ZF and MMSE systems with delay, $N_t = U = 4$, $f_d T_s = 0.04$.

are chosen to be the normalized columns of the matrix [113]

$$\hat{\mathbf{H}}^*[n]\left(\hat{\mathbf{H}}[n]\hat{\mathbf{H}}^*[n] + \frac{U}{P}\mathbf{I}\right)^{-1}. \tag{2.33}$$

From this, we see that the MMSE precoders converge to ZF precoders at high SNR. Therefore, our derivations for the ZF system also apply to the MMSE system at high SNR.

In Figure 2.7, we compare the performance of ZF and MMSE precoding systems with delay[8]. We see that the MMSE precoding outperforms ZF at low to medium SNRs, and converges to ZF at high SNR while converges to BF at low SNR. In addition, it has the same rate ceiling as the ZF system, and crosses the BF curve roughly at the same point, after which we need to switch to the SU mode. Based on this, we can use the second predicted mode switching point (the one at higher SNR) of the ZF system for the MMSE system. We compare the simulation results and calculation

[8]This can also be done in the system with both delay and quantization, which is more time-consuming. As shown in *Lemma 1*, the effects of delay and quantization are equivalent, so the conclusion will be the same.

Table 2.2: Mode Switching Points

	$f_dT_s = 0.03$	$f_dT_s = 0.04$	$f_dT_s = 0.05$
MMSE (Simulation)	44.2 dB	35.7 dB	29.5 dB
ZF (Simulation)	44.2 dB	35.4 dB	28.6 dB
ZF (Calculation)	41.6 dB	32.9 dB	26.1 dB

results by (2.19) and (2.32) for the mode switching points in Table 2.2. For the ZF system, it is the second switching point; for the MMSE system, it is the only switching point. We see that the switching points for MMSE and ZF systems are very close, and the calculated ones are roughly $2.5 \sim 3$ dB lower.

2.7 Conclusion and Future Work

In this chapter, we compare the SU and MU-MIMO transmissions in the broadcast channel with delayed and quantized CSIT, where the amount of delay and the number of feedback bits per user are fixed. The throughput of MU-MIMO saturates at high SNR due to residual inter-user interference, for which a SU/MU mode switching algorithm is proposed. We derive accurate closed-form approximations for the ergodic rates for both SU and MU modes, which are then used to calculate the mode switching points. It is shown that the MU mode is only possible to be active in the medium SNR regime, with a small normalized Doppler frequency and a large codebook size.

In this study, we assume the transmitter knows perfectly the actual received SINR at each active user. In practice, there will inevitably be errors in such information due to estimation error and feedback delay, which will result in rate mismatch, i.e. the transmission rate based on the estimated SINR does not match the actual SINR on the channel, so there will be outage events. How to deal with such rate mismatch is of practical importance and we mention several possible approaches as follows. The full investigation of this issue is left to future work. Considering the outage events, the transmission strategy can be designed based on the actual information symbols successfully delivered to the receiver, denoted as *goodput* in [93, 154]. With

the estimated SINR, another approach is to back off on the transmission rate based on the variance of the estimation error, as did in [141, 142] for the single-antenna opportunistic scheduling system and in [140] for the multiple-antenna opportunistic beamforming system. Combined with user selection, the transmission rate can also be determined based on some lower bound of the actual SINR to make sure that no outage occurs, as did in [90] for the limited feedback system.

For other future work, the MU-MIMO mode studied in this chapter is designed with zero-forcing criterion, which is shown to be sensitive to CSI imperfections, so robust precoding design is needed and the impact of the imperfect CSIT on non-linear precoding should be investigated. As power control is an effective way to combat interference, it is interesting to consider the efficient power control algorithm rather than equal power allocation to improve the performance, especially in the heterogeneous scenario. It is also of practical importance to investigate possible approaches to improve the quality of the available CSIT with a fixed codebook size, e.g. through channel prediction. For practical applications, the impact of more realistic channel models should also be investigated, such as channel correlation. Recently, the comparison of SU-MIMO and MU-MIMO with multiple receive antennas and considering feedback delay was done in [165], where block diagonalization was employed as the MU-MIMO precoder. The investigation of other precoders and more realistic channel models is of interest.

2.8 Appendix

2.8.1 Useful Results for Rate Analysis

In this Appendix, we present some useful results that are used for rate analysis in this chapter. The following lemma will be used frequently in the derivation of the achievable rate.

Lemma 2. *For a random variable X with probability distribution function (pdf) $f_X(x)$ and cumulative distribution function (cdf) $F_X(x)$, we have*

$$\mathbb{E}_X\left[\ln(1+X)\right] = \int_0^\infty \frac{1-F_X(x)}{1+x}dx. \tag{2.34}$$

Proof. The proof follows the integration by parts. □

The following lemma provides some useful integrals for rate analysis, which can be derived using the results in [58].

Lemma 3.

$$I_1(a,b,m) = \int_0^\infty \frac{x^m e^{-ax}}{x+b} dx = \sum_{k=1}^{m} (k-1)!(-b)^{m-k} a^{-k} - (-1)^{m-1} b^m e^{ab} E_1(ab) \tag{2.35}$$

$$\begin{aligned} I_2(a,b,m) &= \int_0^\infty \frac{e^{-ax}}{(x+b)^m} dx \\ &= \begin{cases} e^{ab}E_1(ab) & m=1 \\ \sum_{k=1}^{m-1} \frac{(k-1)!}{(m-1)!} \frac{(-a)^{m-k-1}}{b^k} + \frac{(-a)^{m-1}}{(m-1)!} e^{ab} E_1(ab) & m \geq 2 \end{cases} \end{aligned} \tag{2.36}$$

$$\begin{aligned} I_3(a,b,m) &= \int_0^\infty \frac{e^{-ax}}{(x+b)^m (x+1)} dx \\ &= \sum_{i=1}^{m} (-1)^{i-1} (1-b)^{-i} \cdot I_2\left(a, b, m-i+1\right) + (b-1)^{-m} \cdot I_2\left(a,1,1\right), \end{aligned} \tag{2.37}$$

where $E_1(x)$ is the exponential-integral function of the first order.

2.8.2 Proof of Theorem 1

The average SNR is

$$\begin{aligned} \overline{\mathrm{SNR}}_{BF}^{(QD)} &= \mathbb{E}\left[P \left| \mathbf{h}^*[n] \mathbf{f}^{(QD)}[n] \right|^2 \right] \\ &= P\mathbb{E}\left[\left| (\rho \mathbf{h}[n-1] + \mathbf{e}[n])^* \hat{\mathbf{h}}[n-1] \right|^2 \right] \\ &\overset{(a)}{=} P\mathbb{E}\left[|\rho \mathbf{h}^*[n-1] \hat{\mathbf{h}}[n-1]|^2 \right] + P\mathbb{E}\left[|\mathbf{e}^*[n] \hat{\mathbf{h}}[n-1]|^2 \right] \\ &\overset{(b)}{\leq} P N_t \rho^2 \left(1 - \frac{N_t - 1}{N_t} 2^{-\frac{B}{N_t - 1}} \right) + P\mathbb{E}\left[|\hat{\mathbf{h}}^*[n-1] \cdot [\mathbf{e}[n] \mathbf{e}^*[n]] \cdot \hat{\mathbf{h}}[n-1]| \right] \\ &\overset{(c)}{=} P N_t \rho^2 \left(1 - \frac{N_t - 1}{N_t} 2^{-\frac{B}{N_t - 1}} \right) + P\epsilon_e^2, \end{aligned}$$

As $\mathbf{e}[n]$ is independent of $\mathbf{h}[n-1]$, it is also independent of $\hat{\mathbf{h}}[n-1]$, which gives (a). Step (b) follows (2.11). Step (c) is from the fact $\mathbf{e}[n] \sim \mathcal{CN}(\mathbf{0}, \epsilon_e^2 \mathbf{I}_{N_t})$ and $|\hat{\mathbf{h}}[n-1]|^2 = 1$.

2.8.3 Proof of Theorem 2

Denote $y_1 = \|\mathbf{h}[n-1]\|^2$ and $y_2 = \frac{1}{\epsilon_e^2}|\mathbf{e}^*[n]\tilde{\mathbf{h}}[n-1]|^2$, then $y_1 \sim \chi^2_{2N_t}$, $y_2 \sim \chi^2_2$, and they are independent. The received SNR can be written as $x = \eta_1 y_1 + \eta_2 y_2$, where $\eta_1 = P\rho^2$ and $\eta_2 = P\epsilon_e^2$. The cdf of X is given as [129]

$$F_X(x) = 1 - \left(\frac{\eta_2}{\eta_2 - \eta_1}\right)^{N_t} e^{-x/\eta_2} + e^{-x/\eta_1} \left(\frac{\eta_1}{\eta_2 - \eta_1}\right) \cdot \sum_{i=0}^{N_t-1} \sum_{l=0}^{i} \frac{1}{(i-l)!} \left(\frac{\eta_2}{\eta_2 - \eta_1}\right)^{N_t-1-i} \left(\frac{x}{\eta_1}\right)^{i-l}. \quad (2.38)$$

Denote $a_0 = \frac{\eta_2}{\eta_2 - \eta_1}$ and following *Lemma 2* we have

$$\begin{aligned} &\mathbb{E}_X\left[\ln(1+X)\right] = \int_0^\infty \frac{1 - F_X(x)}{1+x} dx \\ =&a_0^{N_t} \int_0^\infty \frac{e^{x/\eta_2}}{1+x} dx - (1-a_0) \sum_{i=0}^{N_t-1} \sum_{l=0}^{i} \frac{a_0^{N_t-1-i}}{(i-l)!} \left(\frac{1}{\eta_1}\right)^{i-l} \int_0^\infty \frac{x^{i-l} e^{-x/\eta_1}}{1+x} dx \\ =&a_0^{N_t} I_2(1/\eta_2, 1, 1) - (1-a_0) \sum_{i=0}^{N_t-1} \sum_{l=0}^{i} \frac{a_0^{N_t-1-i}}{(i-l)!} \left(\frac{1}{\eta_1}\right)^{i-l} I_1(1/\eta_1, 1, i-l). \end{aligned} \quad (2.39)$$

where $I_1(\cdot,\cdot,\cdot)$ and $I_2(\cdot,\cdot,\cdot)$ are given in (2.35) and (2.36), respectively.

2.8.4 Proof of Theorem 3

The SINR degradation for the ZF system with both channel quantization and delay is

$$\begin{aligned}
\Delta_{ZF,u}^{(QD)} &= \mathbb{E}\left[1+\frac{\gamma}{U}\sum_{u'\neq u}|\mathbf{h}_u^*[n]\mathbf{w}_{u'}^{(QD)}[n]|^2\right] \\
&= 1+\frac{\gamma}{U}\mathbb{E}\left[\sum_{u'\neq u}\left|\left[\rho_u\mathbf{h}_u[n-D]+\mathbf{e}_u[n]\right]^*\mathbf{w}_{u'}^{(QD)}[n]\right|^2\right] \\
&= 1+\frac{\gamma}{U}\rho_u^2\mathbb{E}\left[\sum_{u'\neq u}|\mathbf{h}_u^*[n-D]\mathbf{w}_{u'}^{(QD)}[n]|^2\right]+\frac{\gamma}{U}\mathbb{E}\left[\sum_{u'\neq u}|\mathbf{e}_u^*[n]\mathbf{w}_{u'}^{(QD)}[n]|^2\right] \\
&\overset{(a)}{=} 1+\gamma\rho_u^2 2^{-\frac{B}{N_t-1}}+(U-1)\frac{\gamma}{U}\epsilon_{e,u}^2,
\end{aligned}$$

where (a) follows (2.20) and $|\mathbf{e}_u^*[n]\mathbf{w}_{u'}^{(QD)}[n]|^2 \sim \chi_2^2$.

2.8.5 Proof of Lemma 1

Let $x = \|\mathbf{h}_u[n-D]\|^2\sin^2\theta \sim \Gamma(M-1,\delta)$, $y \sim \beta(1,M-2)$, and x is independent of y. Then the interference term due to quantization is $Z = XY$. The cdf of Z is

$$\begin{aligned}
P_Z(z) &= \int_0^\infty F_{Y|X}\left(\frac{z}{x}\right)f_X(x)dx \\
&= \int_0^z f_X(x)dx+\int_z^\infty\left(1-\left(1-\frac{z}{x}\right)^{M-2}\right)f_X(x)dx \\
&= \int_0^\infty f_X(x)dx-\int_z^\infty\left(1-\frac{z}{x}\right)^{M-2}x^{M-2}\frac{e^{-x/\delta}}{(M-2)!\delta^{M-1}}dx \\
&= 1-e^{-z/\delta}\int_z^\infty(x-z)^{M-2}\frac{e^{-(x-z)/\delta}}{(M-2)!\delta^{M-1}}dx \\
&\overset{(a)}{=} 1-e^{-z/\delta},
\end{aligned} \tag{2.40}$$

where step (a) follows the equality $\int_0^\infty y^M e^{-\alpha y} = M!\alpha^{-(M+1)}$.

2.8.6 Proof of Theorem 4

Assuming each interference term in (2.27) is independent of each other and independent of the signal power term, denote $\sum_{u'\neq u}\rho_u^2|\mathbf{h}_u^*[n-1]\mathbf{f}_{u'}^{(QD)}[n]|^2 = \rho_u^2\delta y_1$

and $\sum_{u' \neq u} |\mathbf{e}_u^*[n]\mathbf{f}_{u'}^{(QD)}[n]|^2 = \epsilon_{e,u}^2 y_2$, then from *Lemma 1* we have $y_1 \sim \chi^2_{2(N_t-1)}$, and $y_2 \sim \chi^2_{2(N_t-1)}$ as $\mathbf{e}_u[n]$ is complex Gaussian with variance $\epsilon_{e,u}^2$ and independent of the normalized vector $\mathbf{f}_{u'}^{(QD)}[n]$. In addition, the signal power $|\mathbf{h}_u^*[n]\mathbf{f}_u^{(QD)}[n]|^2 \sim \chi_2^2$. Then the received SINR for the u-th user is approximated as

$$\gamma_{ZF,u}^{(QD)} \approx \frac{\alpha z}{1+\beta(\delta_1 y_1 + \delta_2 y_2)} \triangleq x, \tag{2.41}$$

where $\alpha = \beta = \frac{P}{U}$, $\delta_1 = \rho_u^2 \delta$, $\delta_2 = \epsilon_{e,u}^2$, $y_1 \sim \chi^2_{2M}$, $y_1 \sim \chi^2_{2M}$, $M = N_t - 1$, $z \sim \chi_2^2$, and y_1, y_2, z are independent of each other.

Let $y = \delta_1 y_1 + \delta_2 y_2$, then the pdf of y, which is the sum of two independent chi-square random variables, is given as [129]

$$p_Y(y) = e^{-y/\delta_1} \sum_{i=0}^{M-1} a_i^{(1)} y^i + e^{-y/\delta_2} \sum_{i=0}^{M-1} a_i^{(2)} y^i = \sum_{j=1}^{2} \sum_{i=0}^{M-1} e^{-y/\delta_j} a_i^{(j)} y^i, \tag{2.42}$$

where

$$a_i^{(1)} = \frac{1}{\delta_1^{i+1}(M-1)!} \left(\frac{\delta_1}{\delta_1-\delta_2}\right)^M \frac{(2(M-1)-i)!}{i!(M-1-i)!} \left(\frac{\delta_2}{\delta_2-\delta_1}\right)^{M-1-i} \tag{2.43}$$

$$a_i^{(2)} = \frac{1}{\delta_2^{i+1}(M-1)!} \left(\frac{\delta_2}{\delta_2-\delta_1}\right)^M \frac{(2(M-1)-i)!}{i!(M-1-i)!} \left(\frac{\delta_1}{\delta_1-\delta_2}\right)^{M-1-i}. \tag{2.44}$$

The cdf of X is

$$\begin{aligned}
F_X(x) &= P\left(\frac{\alpha z}{1+\beta y} \leq x\right) \\
&= \int_0^\infty F_{Z|Y}\left(\frac{x}{\alpha}(1+\beta y)\right) p_Y(y) dy \\
&= \int_0^\infty \left(1 - e^{-\frac{x}{\alpha}(1+\beta y)}\right) p_Y(y) dy \\
&= 1 - e^{-x/\alpha} \int_0^\infty e^{-\beta x y/\alpha} p_Y(y) dy \\
&= 1 - e^{-x/\alpha} \int_0^\infty \left\{ \sum_{j=1}^{2} \sum_{i=0}^{M-1} \exp\left[-\left(\frac{\beta}{\alpha}x + \frac{1}{\delta_j}\right) y\right] a_i^{(j)} y^i \right\} dy \\
&\overset{(a)}{=} 1 - e^{-x/\alpha} \sum_{j=1}^{2} \sum_{i=0}^{M-1} \left[\frac{a_i^{(j)} i!}{\left(\frac{\beta}{\alpha}x + \frac{1}{\delta_j}\right)^{i+1}} \right],
\end{aligned} \tag{2.45}$$

where step (a) follows the equality $\int_0^\infty y^M e^{-\alpha y} = M!\alpha^{-(M+1)}$.

Then the ergodic achievable rate for the u-th user is approximated as

$$\begin{aligned}
R_{ZF,u}^{(QD)} &= \mathbb{E}_\gamma \left[\log_2\left(1+\gamma_{ZF,u}^{(QD)}\right)\right] \\
&\approx \log_2(e)\mathbb{E}_X\left[\ln(1+X)\right] \\
&\overset{(a)}{=} \log_2(e)\int_0^\infty \frac{1-F_X(x)}{x+1}dx \\
&= \log_2(e)\int_0^\infty \sum_{i=0}^{M-1}\sum_{j=1}^{2}\left[a_i^{(j)} i!\left(\frac{\alpha}{\beta}\right)\frac{e^{-x/\alpha}}{\left(x+\frac{\alpha}{\beta\delta_j}\right)^{i+1}(x+1)}\right]dx \\
&\overset{(b)}{=} \log_2(e)\sum_{i=0}^{M-1}\sum_{j=1}^{2}\left[a_i^{(j)} i!\left(\frac{\alpha}{\beta}\right)^{i+1} I_3\left(\frac{1}{\alpha},\frac{\alpha}{\beta\delta_j},i+1\right)\right],
\end{aligned} \tag{2.46}$$

where step (a) follows from *Lemma 2*, step (b) follows the expression of $I_3(\cdot,\cdot,\cdot)$ in (2.37). For equal power allocation, $\alpha=\beta=\frac{P}{U}$, and the expression can be simplified into (2.30).

Chapter 3

Multi-mode Transmission for the MIMO Broadcast Channel

In Chapter 2, it was shown that with imperfect CSIT the MU-MIMO mode can only be activated when the delay is small and the codebook size is large. However, only the MU mode with the number of users $U = N_t$ is considered. In this chapter, we consider a more general scenario where the number of active users, called the *transmission mode*, can vary from 1 to N_t, and the mode is selected adaptively to maximize the sum throughput. The problem we explore in this chapter is: *How many users should be served to maximize the sum throughput in the MIMO broadcast channel with imperfect CSIT?*

3.1 Related Work

With a fixed codebook size, or other fixed CSIT imperfections such as delay, the throughput of limited feedback MU-MIMO systems saturates at high SNR [43, 75, 164, 169]. An approach to solve this problem is adaptive switching between the SU and MU modes, as the SU mode does not suffer from residual interference at high SNR.

SU/MU mode switching algorithms for the random beamforming system were proposed in [14, 91], where each user feeds back its preferred mode and the channel quality information (CQI). Random beamforming is a low-complexity linear precoding MU-MIMO transmission technique, which employs random orthogonal beamforming vectors and selects users based on the feedback of the preferred beam index and the associated signal-to-interference-plus-noise ratio (SINR) from each user [124].

Another practical MU-MIMO transmission technique is linear precoding com-

bined with quantized channel information feedback [75, 156]. Recently, there have been some studies on scalar CQI feedback for MU-MIMO with zero-forcing (ZF) precoding and limited feedback, which achieves full multi-user diversity gain [90, 137, 156]. Such CQI contains information about both the signal and interference power, based on which the BS can determine the number of active users that should be scheduled.

For the MIMO downlink with the number of receive antennas greater than or equal to the number of transmit antennas, an adaptive SU/MU mode switching algorithm was proposed in [94], which also considered correlation at transmit antennas. The switching is based on partial CSIT, including spatial correlation factor, the number of transmit and receive antennas, and the number of users. Then each user feeds back the information about its achievable rate, based on which scheduling is performed.

Most prior mode switching algorithms are based on instantaneous CSIT [14, 90, 91, 137]. Another approach is to switch based on the statistics of the channel. In [39], the number of active users for a ZF precoding system was optimized through asymptotic analysis to maximize the spectral efficiency, based on average SNR and codebook size. Chapter 2 also proposed an SU/MU switching algorithm for ZF precoding with delayed and quantized CSIT. The mode switching is based on average SNR, normalized Doppler frequency, and codebook size, which are computable at the BS.

With a large number of users, scheduling is required to select a subset of users to serve at each instant, which is based on the available CSIT. It provides multi-user diversity gain [88, 147] at the expense of feedback of instantaneous channel information from many users. For MIMO systems, the amount of total feedback grows with both the number of users and the number of antennas, which is undesirable. This amount can be reduced with the threshold-based feedback strategy [49, 60, 68]. In this chapter, we approach this problem from a different perspective, and propose scheduling algorithms based on the statistics of the channel information.

Both switching algorithms in [39] and Chapter 2 require only the selected users to feed back their instantaneous channel information, which is desirable for systems with a large number of users. The technique in Chapter 2 is based on non-asymptotic analysis, and thus it can better characterize the system behavior with different parameters, e.g. the operating regions for both SU and MU modes can be determined for different delays or different codebook sizes. But it only switches between the SU mode and the MU mode serving as many users as transmit antennas. Furthermore, scheduling among a large number of users has not been considered therein. In this chapter, we propose a more general multi-mode switching algorithm that adaptively selects different modes for different scenarios, and propose user scheduling algorithms based on this.

3.2 Contributions

In this chapter, we consider a MIMO broadcast channel with delayed and quantized CSIT. The main contributions are as follows.

Throughput Analysis and Mode Selection. First we derive accurate closed-form approximations for the achievable rates for different modes, based on which the mode that provides the highest throughput can be selected. Such multi-mode transmission improves the spectral efficiency by balancing transmit array gain, spatial division multiplexing gain, and residual interference. It is shown that the full MU mode, which serves as many users as the number of transmit antennas, normally will not be activated, as it has the highest residual interference and no array gain. It is also shown that with multi-mode transmission ZF precoding provides performance close to MMSE precoding.

User Scheduling Algorithms. Next we propose user scheduling algorithms based on slow time-varying channel information, including the average SNR and normalized Doppler frequency. Only the scheduled users need to feed back their instantaneous channel information for the precoder design, reducing the amount of feedback. In the homogeneous network, it is shown that the proposed scheduling algorithm pro-

vides a throughput gain over both the SU-MIMO system and the dual-mode switching algorithm proposed in Chapter 2 at medium SNR. In the heterogeneous network where users experience different path losses, the multi-mode based scheduling algorithm is able to provide throughput close to the one based on instantaneous feedback from all the users.

3.3 System Model

We consider a MIMO broadcast channel with N_t antennas at the transmitter and U single-antenna mobiles. The discrete-time complex baseband received signal at the u-th user at time n is given as

$$y_u[n] = \mathbf{h}_u^*[n] \sum_{v=1}^{M} \mathbf{f}_v[n] x_v[n] + z_u[n], \tag{3.1}$$

where M is the active mode, i.e. the number of active users, $M = 1, 2, \cdots, N_t$, $\mathbf{h}_u[n]$ is the channel vector for the u-th user, $\mathbf{h}_u[n] \in \mathbb{C}^{(N_t \times 1)}$, and $z_u[n]$ is normalized complex additive Gaussian noise, $z_u[n] \sim \mathcal{CN}(0, 1)$. $x_u[n]$ and $\mathbf{f}_u[n]$ are the transmit signal and precoding vector for the u-th user, and $\mathbf{f}_u[n] \in \mathbb{C}^{(N_t \times 1)}$. The transmit power constraint is $\mathbb{E}\left[\sum_{u=1}^{M} |x_u[n]|^2\right] = P$, and we assume equal power allocation among different users. As the noise is normalized, P is also the average SNR.

As in Chapter 2, eigen-beamforming is applied for the SU mode ($M = 1$), which transmits along the channel direction and is optimal for SU-MIMO with perfect CSIT [136]. ZF precoding is used for the MU mode ($1 < M \leq N_t$), as it is possible to derive closed-form results due to its simple structure, and it is optimal among the set of all linear precoders at asymptotically high SNR [74].

To assist the analysis, we assume that the channel $\mathbf{h}_u[n]$ is well modeled as a spatially white Gaussian channel, with entries $h_i[n] \sim \mathcal{CN}(0, 1)$. We assume perfect channel state information (CSI) at the receiver, while the transmitter obtains CSI through limited feedback from the receiver. In addition, there is delay in the available CSIT. The models of the CSI delay and limited feedback are the same as used in Chapter 2, presented in Section 2.3.1 and Section 2.3.2, respectively.

3.4 Throughput Analysis and Mode Selection

In this section, we investigate the impact of imperfect CSIT on the MIMO broadcast channel and derive achievable rates for different transmission modes. It is shown that the number of active users is closely related to transmit array gain, spatial division multiplexing gain, and residual inter-user interference. Multi-mode transmission is proposed to adaptively select the active mode to balance between these effects and maximize the throughput. In this section, we consider a homogeneous scenario, where each user has i.i.d. channel, the same delay and quantization codebook size. The BS determines the active mode, i.e. the number of active users to serve. Detailed scheduling algorithms for different scenarios will be proposed in Section 3.5.

3.4.1 Perfect CSIT

We first consider the system with perfect CSIT, which serves as the basis for the analysis of the impact of imperfect CSIT.

SU-MIMO (Eigen-beamforming), $M = 1$ With perfect CSIT, the beamforming (BF) vector is the channel direction, i.e. $\mathbf{f}[n] = \tilde{\mathbf{h}}[n]$. The ergodic rate is the same as that of a maximal ratio combining diversity system, given in [10] as

$$\begin{aligned} R_{CSIT}(1) = R_{BF}(\gamma, N_t) &\triangleq \mathbb{E}_{\mathbf{h}}\left[\log_2\left(1+\gamma|\mathbf{h}^*[n]\mathbf{f}[n]|^2\right)\right] \\ &= \log_2(e)e^{1/\gamma}\sum_{k=0}^{N_t-1}\frac{\Gamma(-k,1/\gamma)}{\gamma^k}, \end{aligned} \tag{3.2}$$

where $\Gamma(\alpha, x) = \int_x^\infty t^{\alpha-1}e^{-t}dt$ is the complementary incomplete gamma function, and $R_{BF}(\gamma, n)$ is the rate function for the diversity system with SNR γ and diversity order n. The BF system provides transmit array gain[1] N_t as $\mathbb{E}_{\mathbf{h}}\left[\gamma|\mathbf{h}^*[n]\mathbf{f}[n]|^2\right] = N_t\gamma$.

[1]The array gain is defined as the increase in the average combined SNR over the average SNR on each branch [56].

MU-MIMO (Zero-forcing), $1 < M \leq N_t$ The received SINR for the u-th user in a linear precoding MU-MIMO system in mode M is given by

$$\mathrm{SINR}_u(M) = \frac{\frac{\gamma}{M}|\mathbf{h}_u^*[n]\mathbf{f}_u[n]|^2}{1 + \frac{\gamma}{M}\sum_{v\neq u}|\mathbf{h}_u^*[n]\mathbf{f}_v[n]|^2}. \tag{3.3}$$

Denote $\mathbf{H}[n] = [\mathbf{h}_1[n], \mathbf{h}_2[n], \cdots, \mathbf{h}_U[n]]^*$, and the pseudo-inverse of $\mathbf{H}[n]$ as $\mathbf{F}[n] = \mathbf{H}^\dagger[n] = \mathbf{H}^*[n](\mathbf{H}[n]\mathbf{H}^*[n])^{-1}$. The ZF precoding vector for the u-th user is obtained by normalizing the u-th column of $\mathbf{F}[n]$. Therefore, $\mathbf{h}_u^*[n]\mathbf{f}_v[n] = 0$, $\forall u \neq v$, i.e. there is no inter-user interference and each user gets an equivalent interference-free channel. The SINR for the u-th user becomes

$$\mathrm{SINR}_{ZF,u}(M) = \frac{\gamma}{M}|\mathbf{h}_u^*[n]\mathbf{f}_u[n]|^2. \tag{3.4}$$

Due to the isotropic nature of i.i.d. Rayleigh fading, such orthogonality constraints to precancel inter-user interference consume $M-1$ degrees of freedom at the transmitter. As a result, the effective channel gain of each parallel channel is a chi-square random variable with $2(N_t-M+1)$ degrees of freedom [76, 96], i.e. $|\mathbf{h}_u^*[n]\mathbf{f}_u[n]|^2 \sim \chi^2_{2(N_t-M+1)}$. Therefore, the channel for each user is equivalent to a diversity channel with order $(N_t - M + 1)$ and effective SNR $\frac{\gamma}{M}$. The achievable ergodic rate for the u-th user in mode M is

$$R_{CSIT,u}(M) = R_{BF}\left(\frac{\gamma}{M}, N_t - M + 1\right), \tag{3.5}$$

The achievable ergodic sum rate for the ZF system of mode M is

$$R_{CSIT}(M) = \sum_{u=1}^{M} R_{CSIT,u}(M) \overset{(a)}{=} MR_{BF}\left(\frac{\gamma}{M}, N_t - M + 1\right). \tag{3.6}$$

The equality (a) follows the homogeneous nature of the network. When $M = 1$, this reduces to (3.2).

Mode Selection From (3.6), the system in mode M provides a spatial division multiplexing gain of M and an array gain of $(N_t - M + 1)$ for each user. As M increases, the achievable spatial division multiplexing gain increases but the array

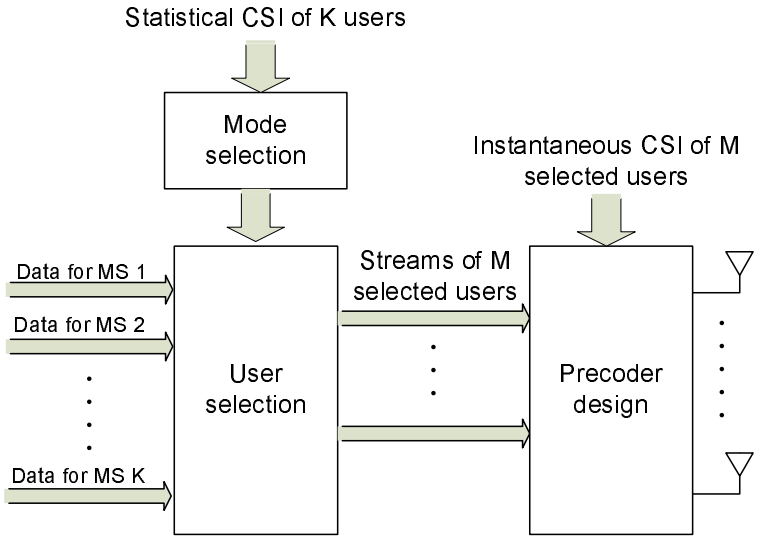


Figure 3.1: The illustration of multi-mode transmission.

gain decreases. Therefore, there is a tradeoff between the achievable array gain and the spatial division multiplexing gain. From (3.6), the mode that achieves the highest throughput for the given average SNR can be determined as

$$M^{\star} = \arg \max_{1 \leq M \leq N_t} R_{CSIT}(M). \tag{3.7}$$

Note that this is a very simple optimization problem, as only N_t values need to be computed and compared. This transmission strategy that adaptively adjusts the number of active users is denoted as *multi-mode transmission*. The illustration of multi-mode transmission is shown in Figure 3.1.

The numerical results for multi-mode transmission with perfect CSIT are plotted in Figure 3.2. We see that in the low SNR regime lower modes provide higher throughputs than higher modes due to higher array gains, while in the high SNR regime higher modes provide higher rates with larger spatial division multiplexing gains. Switching between multiple modes efficiently makes the tradeoff between the array gain and the spatial division multiplexing gain.

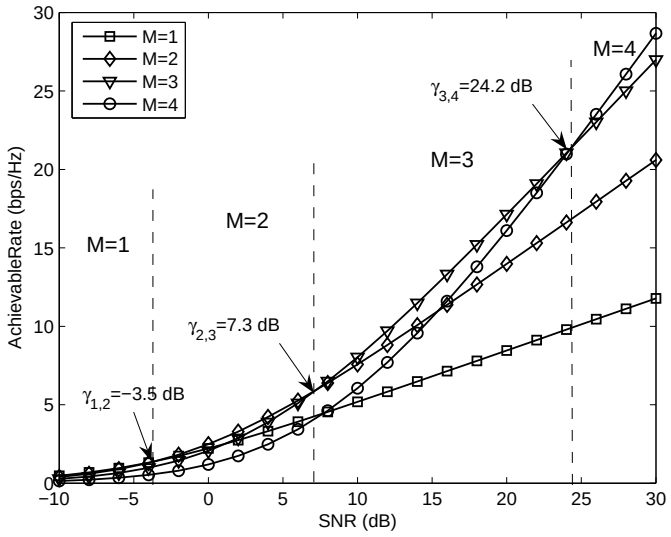


Figure 3.2: The numerical results for the multi-mode transmission with perfect CSIT, $U = N_t = 4$.

3.4.2 Imperfect CSIT

In this section, we consider imperfect CSIT, including both delay and channel quantization. As it is difficult to derive the exact achievable rate for such a system, we provide accurate closed-form approximations for mode selection. The achievable ergodic rate for the SU mode $M = 1$ in such a system is provided in Section 2.5.1.

For the MU mode with delay and quantization, the precoding vectors are designed based on the quantized channel directions with delay, which achieve $\hat{\mathbf{h}}_u^*[n-1]\mathbf{f}_v^{(QD)}[n] = 0$, $\forall u \neq v$. The SINR for the u-th user in mode M is given as

$$\gamma_{ZF,u}^{(QD)}(M) = \frac{\frac{P}{M}|\mathbf{h}_u^*[n]\mathbf{f}_u^{(QD)}[n]|^2}{1 + \frac{P}{M}\sum_{v\neq u}|\mathbf{h}_u^*[n]\mathbf{f}_v^{(QD)}[n]|^2}. \tag{3.8}$$

We assume the mobile users can perfectly estimate the noise and interference and feed back these information to the transmitter, so the achievable rate for the u-th user is given as

$$R_{QD,u}(M) = \mathbb{E}_\gamma\left[\log_2(1+\gamma_{ZF,u}^{(QD)}(M))\right]. \tag{3.9}$$

The same metric is used in [43, 75].

Based on the results in [21, 75], we can derive an upper bound for the rate loss for the u-th user due to imperfect CSIT, stated in the following theorem.

Theorem 5. *The rate loss for the u-th user with imperfect CSIT compared to that with perfect CSIT is upper bounded by*

$$R_{CSIT,u} - R_{QD,u} \leq \log_2 \Delta_u^{(QD)}, \tag{3.10}$$

where $\Delta_u^{(QD)}$ *is the average noise plus residual interference, given by*

$$\begin{aligned}\Delta_u^{(QD)} &= \mathbb{E}\left[1 + \frac{P}{M}\sum_{v\neq u} |\mathbf{h}_u^*[n]\mathbf{f}_v^{(QD)}[n]|^2\right] \\ &= 1 + \left(1 - \frac{1}{M}\right) P\left(\rho_u^2 \frac{N_t}{N_t - 1} 2^{-\frac{B}{N_t - 1}} + \epsilon_{e,u}^2\right).\end{aligned} \tag{3.11}$$

Proof. The average noise plus interference $\Delta_u^{(QD)}$ can be derived in the similar way as for *Theorem 3* in Chapter 2. The upper bound is obtained by following the approach in [21, 75]. □

Remark 6. *From* (3.11)*, we see that residual interference/rate loss depends on delay, codebook size,* N_t*, and* M*. It increases with delay, and decreases with codebook size. It also increases with* P*, which makes the system interference-limited at high SNR. With other parameters fixed, the residual interference increases as* M *increases, which means it may not be desirable to serve too many users.*

The bound analysis in (3.10) provides helpful insights on the effects of system parameters on the rate loss, but it is not accurate enough for mode selection. To get a better approximation for the achievable rate, we first approximate the signal term as

$$\begin{aligned}P_S &= \frac{P}{M}|\mathbf{h}_u^*[n]\mathbf{f}_u^{(QD)}[n]|^2 \\ &= \frac{P}{M}|(\rho_u \mathbf{h}_u[n-1] + \mathbf{e}_u[n])^* \mathbf{f}_u^{(QD)}[n]|^2 \\ &\overset{(a)}{\approx} \frac{P}{M}|\rho_u \mathbf{h}_u^*[n-1]\mathbf{f}_u^{(QD)}[n]|^2 \\ &\overset{(b)}{\approx} \frac{P}{M}|\rho_u|\|\mathbf{h}_u[n-1]\|\hat{\mathbf{h}}_u^*[n-1]\mathbf{f}_u^{(QD)}[n]|^2,\end{aligned} \tag{3.12}$$

where step (a) removes $\mathbf{e}_u^*[n]\mathbf{f}_u^{(QD)}[n]$, which is normally very small. Step (b) approximates the actual channel direction by the quantized version, which is justified for small quantization error[2]. As the quantized channel direction $\hat{\mathbf{h}}_u[n-1]$ is independent of each other, and $\mathbf{f}_u^{(QD)}[n]$ is designed to lie in the nullspace of $\hat{\mathbf{h}}_v[n-1]$, $\forall v \neq u$, similar to the case of perfect CSIT, $\|\mathbf{h}_u[n-1]\| \cdot |\hat{\mathbf{h}}_u^*[n-1]\mathbf{f}_u^{(QD)}[n]|^2 \sim \chi^2_{2(N_t-M+1)}$. So it also provides an array gain of $N_t - M + 1$ for each user.

The received SINR for the u-th user can then be approximated as

$$\gamma_{ZF,u}^{(QD)}(M) \approx \frac{\frac{P}{M}\rho_u^2\|\mathbf{h}_u[n-1]\| \cdot |\hat{\mathbf{h}}_u^*[n-1]\mathbf{f}^{(QD)}[n]|^2}{1+\frac{P}{M}\sum_{v\neq u}\rho_u^2|\mathbf{h}_u^*[n-1]\mathbf{f}_v^{(QD)}[n]|^2+\frac{P}{M}\sum_{v\neq u}|\mathbf{e}_u^*[n]\mathbf{f}_v^{(QD)}[n]|^2}. \quad (3.13)$$

The approximation for the denominator comes from removing the terms with both $\mathbf{e}_u[n]$ and $\mathbf{f}_v^{(QD)}[n]$. As shown by *Lemma 1* in Chapter 2, the interference term due to quantization, $|\mathbf{h}_u^*[n-1]\mathbf{f}_v^{(QD)}[n]|^2$, can be well approximated as an exponential random variable with mean $\delta = 2^{-\frac{B}{N_t-1}}$. The interference term due to delay, $|\mathbf{e}_u^*[n]\mathbf{f}_v^{(QD)}[n]|^2$, is also an exponential random variable with mean $\epsilon_{e,u}^2$ as $\mathbf{e}_u[n] \sim \mathcal{CN}(\mathbf{0}, \epsilon_{e,u}^2\mathbf{I})$, $|\mathbf{f}_v[n]|^2 = 1$. Based on these, the approximation for the achievable rate for the MU mode can be derived, given in the following theorem.

Theorem 6. *The achievable ergodic rate for the u-th user in the MU system of mode M ($M > 1$) with both delay and channel quantization can be approximated by*

$$R_{QD,u}(M) \approx \log_2(e) \sum_{i=0}^{N_t-M}\sum_{j=1}^{2}\sum_{k=0}^{M-2}\sum_{l=0}^{i}\frac{a_k^{(j)}(l+k)!}{l!(i-l)!}\alpha^{l+k-i+1} \cdot \hat{I}_3\left(\frac{1}{\alpha}, \frac{\alpha}{\delta_j}, i, l+k+1\right), \quad (3.14)$$

where $\alpha = \frac{\rho_u^2 P}{M}$, $\delta_1 = \frac{\rho_u^2 P\delta}{M}$, $\delta_2 = \frac{\epsilon_{e,u}^2 P}{M}$, $a_i^{(1)}$ and $a_i^{(2)}$ are given in (3.28) *and* (3.29)*, and $\hat{I}_3(\cdot,\cdot,\cdot,\cdot)$ is the integral given in* (3.22) *in Appendix 3.8.1.*

Proof. See Appendix 3.8.2. Note that *Theorem 4* in Chapter 2 is a special case of this theorem. □

[2]This approximation works well for a large value of B, and for small B the same adjustment as used in *lemma 12* in Chapter 5 can be adopted to increase the accuracy

Remark 7. *To calculate* (3.14) *for a given user, we need only information about its correlation coefficient ($\rho_u^2 = 1 - \epsilon_{e,u}^2$), codebook size ($B$), and average SNR ($P$). Such information is normally fixed or changes slowly. Each user can feed back and update its own information, and then the BS can calculate the achievable rate. Instead, each user may also calculate the achievable rate and feeds back the preferred mode index. Note that the calculation and the mode selection is only done when the parameter changes, such as path loss change due to mobility.*

As a special case, for the system with delay only[3], following (3.12) and (3.13) the received SINR for the u-th user is approximated as

$$\gamma_{ZF,u}^{(D)} \approx \frac{\frac{P}{M}|\rho_u \mathbf{h}_u^*[n-1]\mathbf{f}_u[n]|^2}{1 + \frac{P}{M}\sum_{v\neq u}|\mathbf{e}_u^*[n]\mathbf{f}_v[n]|^2}, \tag{3.15}$$

for which the achievable ergodic rate is provided as follows.

Corollary 3. *The achievable ergodic rate for the u-th user in the delayed system of mode M ($M > 1$) can be approximated by*

$$R_{D,u}(M) \approx \log_2(e) \sum_{i=0}^{N_t-M}\sum_{l=0}^{i}\binom{M+l-2}{l}\frac{\alpha^{M-1+l-i}}{\beta^{M-1}(i-l)!}\cdot \hat{I}_3\left(\frac{1}{\alpha}, \frac{\alpha}{\beta}, i, M-1+l\right), \tag{3.16}$$

where $\alpha = \frac{\rho_u^2 P}{M}$, $\beta = \frac{\epsilon_{e,u}^2 P}{M}$, and $\hat{I}_3(\cdot,\cdot,\cdot,\cdot)$ is given in (3.22) *in Appendix 3.8.1.*

Proof. Following the same steps in Appendix 3.8.2 with $\delta_1 = 0$. □

In the following, we provide high SNR approximations for MU modes that can be used to analyze the performance in the interference-limited region.

Theorem 7. *The achievable ergodic rate for the u-th user in the system with both delay and channel quantization in the MU mode $M > 1$ at high SNR is approximated as*

$$R_{QD,u}^h(M) \approx \log_2(e) \sum_{i=0}^{N_t-M}\sum_{j=1}^{2}\sum_{k=0}^{M-2}\frac{a_k^{(j)}\hat{\alpha}^{k+1}(k+i)!}{i!}\hat{I}_6\left(\frac{\hat{\alpha}}{\hat{\delta}_j}, i, k+i+1\right), \tag{3.17}$$

[3]As shown in Chapter 2, with quantization cell approximation, the effects of delay and channel quantization are equivalent, so the results for the delayed system also applies to the limited feedback system.

where $\hat{\alpha} = \rho_u^2$, $a_i^{(1)}$ *and* $a_i^{(2)}$ *are given in* (3.28) *and* (3.29) *with* $\hat{\delta}_1 = \rho_u^2\delta$, $\hat{\delta}_2 = \epsilon_{e,u}^2$, *and* $\hat{I}_6(\cdot,\cdot,\cdot)$ *is the integral given in* (3.25) *in Appendix 3.8.1.*

Proof. See Appendix 3.8.3. □

The high SNR result for the system with delay is provided in the following corollary.

Corollary 4. *The achievable sum rate for the u-th user in the delayed system in MU mode* $M > 1$ *at high SNR is approximated as*

$$R_{D,u}^h(M) \approx \log_2(e) \sum_{i=0}^{N_t-M} \binom{M+i-2}{i} \hat{\alpha}^{M-1} \hat{I}_6(\hat{\alpha}, i, M-1+i), \tag{3.18}$$

where $\hat{\alpha} = \frac{\rho_u^2}{\epsilon_{e,u}^2}$, *and* $\hat{I}_6(\cdot,\cdot,\cdot)$ *is given in* (3.25) *in Appendix 3.8.1.*

Proof. Following the steps in Appendix 3.8.3 with $\delta_1 = 0$. □

Based on (3.14), the active mode that achieves the highest ergodic throughput in the system with both delay and channel quantization is selected according to

$$M^\star = \arg \max_{1 \le M \le N_t} R_{QD}(M), \tag{3.19}$$

where $R_{QD}(M) = \sum_{u=1}^{M} R_{QD,u}(M)$, with $R_{QD,u}(M)$ given in (3.14).

Remark 8. *Considering* (3.11), (3.12), *and* (3.14), *the mode* M *is now related to residual interference, transmit array gain, and spatial division multiplexing gain. The idea of multi-mode transmission is to balance between these effects to maximize the system throughput. The mode selection is based on fixed system parameters – the number of transmit antennas and the codebook size – and slow time-varying channel information – average SNR and normalized Doppler frequency.*

To show the accuracy of the derived approximations, numerical results are provided in Figure 3.3 for different modes, with $N_t = 4$, $B = 18$ bits, $v = 10$ km/hr, and $T_s = 1$ msec. We see that the approximation is very accurate at low to medium

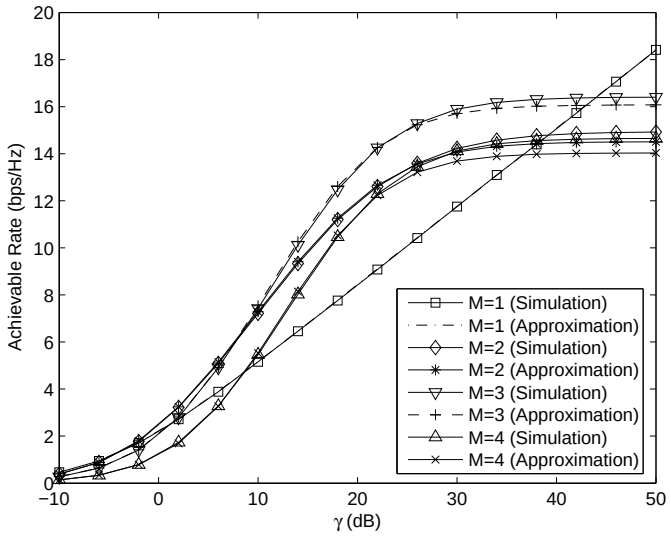


Figure 3.3: Simulation results and approximations for different M, $N_t = 4$, $v = 10$ km/hr, $T_s = 1$ msec, $B = 18$ bits.

SNRs. At high SNR, when the sum rate of the MU mode saturates, the approximation becomes a lower bound, and the accuracy decreases as M increases. Interestingly, we see that the mode $M = 3$ always provides a higher throughput than the full MU mode $M = 4$. This is due to the fact that the full mode has the highest level of residual interference, as shown in (3.11), while it provides no array gain. Therefore, it is desirable to serve fewer than N_t users. We see that the multi-mode transmission is able to provide a throughput gain around 2 bps/Hz over the dual-mode switching in Chapter 2 at medium SNR.

It is easy to verify that the high SNR result (3.17) matches the approximation, and can predict the behavior in the interference-limited region. In Figure 3.4, we plot $M^* = \arg\max_{M>1} R_{QD}^h(M)$ for different normalized Doppler frequency $f_d T_s$, i.e. the mode with the highest sum rate in the interference-limited region. We see that M^* is different for different $f_d T_s$. For $B = 10$ bits, the mode $M = 2$ always has the highest throughput in the considered $f_d T_s$ range, as it provides a higher array gain and has lower residual interference than the higher modes. For $B = 15$ bits, as the CSIT accuracy is improved compared to $B = 10$ bits, $M = 3$ has the highest rate at high

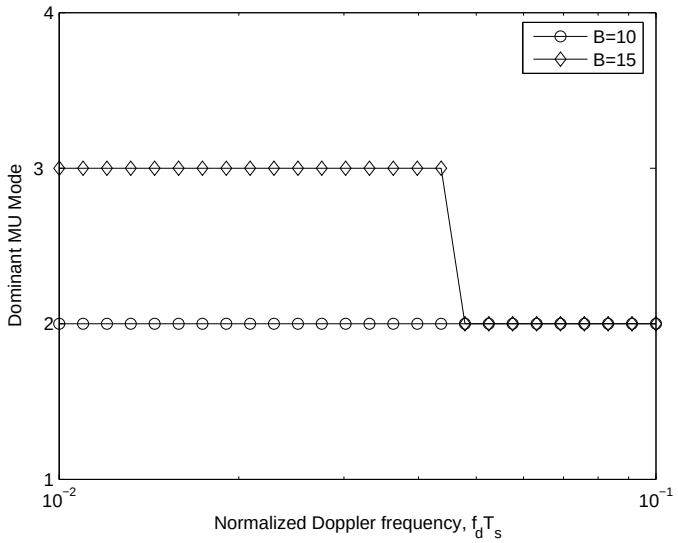


Figure 3.4: The MU mode with the highest rate ceiling for different f_dT_s, with $N_t = 4$.

SNR when f_dT_s is small, but $M = 4$ still has a lower throughput. So with both delay and channel quantization, the highest mode $M = N_t$ is normally not preferable.

3.5 User Scheduling with Imperfect CSIT

For the MIMO broadcast channel with linear precoding, the number of users that can be supported simultaneously is constrained by the number of transmit antennas, so we need to select and transmit to a subset of users in each time slot since typically $U \gg N_t$. With the throughput analysis in Section 3.4, we propose user scheduling algorithms for both homogeneous and heterogeneous networks, which are based not on the instantaneous CSI feedback but on the slow time-varying channel information including average received SNR and the normalized Doppler frequency.

3.5.1 Scheduling in Homogeneous Networks

As shown in Section 3.4, the BS can estimate the achievable ergodic rate for each user based on the information of average SNR, normalized Doppler frequency, and codebook size, which are relatively easy to obtain and are fixed or change slowly.

We propose to perform user scheduling based on such ergodic rate estimation. In this way, the BS can calculate the expected achievable rates for different user combinations, and select users without instantaneous CSI feedback. The proposed algorithms can greatly reduce the amount of CSI feedback, as only the selected M users ($M \ll U$) need to feed back their instantaneous CSI for the precoder design. Such user scheduling is denoted as *Multi-Mode Transmission (MMT) based scheduling*. We first propose the MMT-based scheduling algorithm for the homogeneous network where all the users have the same average SNR, normalized Doppler frequency, and number of feedback bits. We assume the BS has knowledge of the average SNR and the normalized Doppler frequency of each user.

In homogeneous networks, the MMT-based scheduling algorithm determines the number of users that should be served according to (3.19), i.e. determining the active mode $M^\star$. No further selection can be done due to the lack of instantaneous CSIT and the homogeneous nature of the network. Such scheduling is equivalent to *mode selection* considered in Section 3.4. Then $M^\star$ users are selected randomly or with round-robin scheduling from the user pool. Only these $M^\star$ users need to feed back their instantaneous channel information, which is used to design the precoding vectors. Although this algorithm cannot exploit multi-user diversity gain, it will be shown to be able to provide a throughput gain over the SU-MIMO system. In addition, it provides temporal fairness.

3.5.2 Scheduling in Heterogeneous Networks

Most practical networks are heterogeneous, with different average SNRs due to pathloss, mobility and delay that can vary massively from user to user. For such networks, we consider two different scheduling schemes: *round-robin scheduling*, which serves users in order, and *opportunistic scheduling*, which schedules users based on channel conditions including pathloss, delay, and mobility. For the MMT-based round-robin scheduling, the users are served in order, and in each time slot the BS determines how many users to serve, i.e. it performs *mode selection*; For the MMT-based opportunistic scheduling, the BS schedules users that provide the

maximum ergodic sum rate, i.e. it performs both *mode selection* and *user selection*.

MMT-based Round-robin Scheduling The MMT-based scheduling algorithm for the homogeneous network can be modified to apply in the heterogeneous network. In each time slot, the BS adds users from the user queue into the selected subset before the achievable ergodic sum rate begins to decrease. The achievable rate for each user is calculated by (3.14). Then the selected users are informed and feed back their instantaneous CSI for the precoder design.

Remark 9. *The above algorithm is suitable for systems with scheduling independent of the channel status, such as round-robin or the ones based on the queue length. In addition, as each mobile can calculate its own preferred mode $M_u^\star$, it does not need to feed back information about the average SNR and the normalized Doppler frequency, but only needs to report the preferred mode $M_u^\star$ when it changes.*

MMT-based Opportunistic Scheduling For a heterogeneous network, additional selection based on different channel states, i.e. different pathlosses or different delays/mobilities, can be applied to improve the system throughput. The scheduling metric is the achievable ergodic rate, calculated by (3.14). The transmission protocol is as follows.

1. Each user sends update information about its average received SNR and normalized Doppler frequency to the BS.
2. The BS calculates the achievable ergodic rate for different user combinations by (3.14), and selects the one with the highest rate.
3. The BS informs the selected users of the scheduling result.
4. The scheduled users feed back their instantaneous channel information.
5. The BS designs precoding vectors based on the available CSIT, and sends pilot symbols for the users to estimate the received SINR.

6. The scheduled users feed back the instantaneous SINR, assumed to be perfectly fed back to the BS. Then the BS determines the transmission rate and starts the transmission.

Note that step (i) and (ii) only need to be performed when the average SNR/normalized Doppler frequency changes for some user. In step (ii), the user selection algorithm proposed in Appendix 3.8.4 will be applied.

Remark 10. *As only the scheduled users need to feed back instantaneous channel information, the proposed algorithm reduces the feedback amount. It is suitable for the system where there is a constraint on the total feedback bits. Fairness can be considered by employing user selection algorithms such as the proportional fairness algorithm [147].*

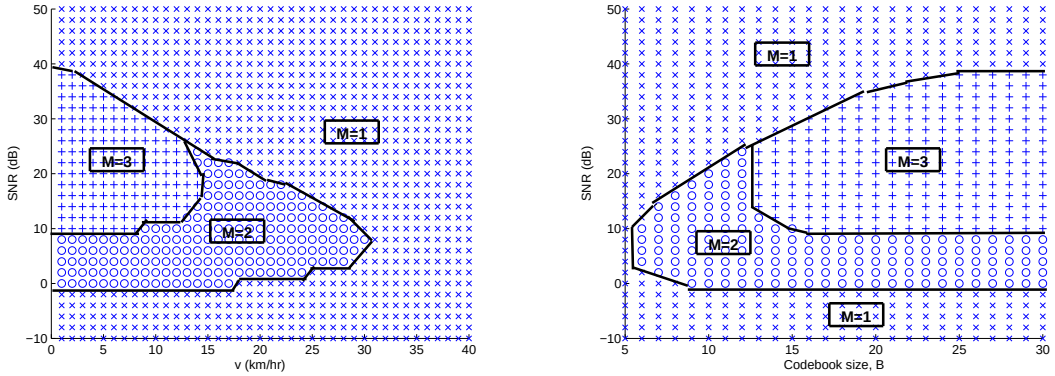
3.6 Numerical Results

This section presents numerical results to demonstrate the performance of our proposed transmission strategies.

3.6.1 Operating Regions for Multi-mode Transmission

For a homogeneous network, based on the results in Section 3.4, the number of active users, i.e. the active mode $M^\star$, can be determined for a given scenario. Accordingly, the operating regions for different modes can be plotted for different system parameters. In this section, we consider parameters used in the 3GPP LTE [2, 3]. The Advanced Wireless Services (AWS) spectrum, which is one of the prime candidates for initial LTE deployment in the US, is considered, i.e. the carrier frequency is $f_c = 2.1$ GHz. In LTE, the minimum size of radio resource that can be allocated in the time domain is one subframe of 1 msec, and we assume the delay in the CSIT is two subframes, i.e. the delay is $\tau = 2$ msec.

Figure 3.5(a) and Figure 3.5(b) show the operating regions for the system with both delay and channel quantization, for different mobility v and different feedback



(a) Different v, $f_c = 2.1$ GHz, $\tau = 2$ msec, $B = 15$ bits.

(b) Different B, $f_c = 2.1$ GHz, $\tau = 2$ msec., and $v = 10$ km/hr.

Figure 3.5: Operating regions for different modes with both CSI delay and channel quantization, $N_t = 4$. The mode $M = i$ means that there are i active users. In this plot, '$\times$' is for $M = 1$, 'o' is for $M = 2$, '+' is for $M = 3$, and '$\square$' is for $M = 4$.

bits B, respectively, where each mark on the figure denotes the type of the active mode. There are several key observations.

1. For the given v and B, the SU mode ($M = 1$) will be active at both low and high SNRs, due to its array gain and the robustness to imperfect CSIT, respectively.

2. For MU modes to be active, v needs to be small while B needs to be large. Specifically, to activate $M = 2$, we need $v \leq 30$ km/hr with $B = 15$ bits as in Figure 3.5(a), and need $B \geq 6$ bits with $v = 10$ km/hr and $T_s = 2$ msec as in Figure 3.5(b); to activate $M = 3$, we need $v \leq 14$ km/hr with $B = 15$ bits and $B \geq 13$ bits with $v = 10$ km/hr. Note that in LTE each user only feeds back 4 bits to indicate its channel direction.

3. The full MU mode $M = N_t$ is not active at all with the considered parameters, as it suffers from the highest residual interference and does not provide array gain.

In Figure 3.6, we compare zero-forcing precoding with multi-mode transmission (ZF-MMT) and the single-user beamforming (BF) transmission, $f_c = 2.1$ GHz, $\tau = 2$ msec., and $v = 10$ km/hr. We see that for $B = 4$ the curves of ZF-MMT

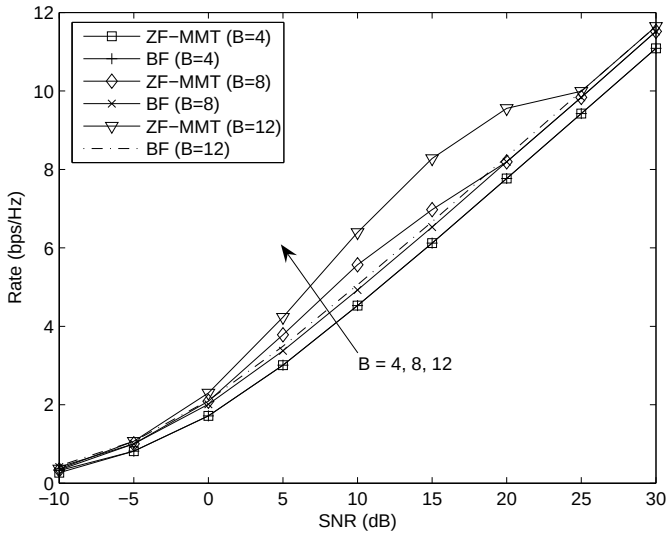


Figure 3.6: Simulation results of multi-mode transmission with different B, $N_t = 4$, $f_c = 2.1$ GHz, $\tau = 2$ msec., and $v = 10$ km/hr.

and BF overlap, which means no MU mode is activated. This confirms the result in Figure 3.5(b). For $B = 8$, ZF-MMT provides throughput gain over BF for SNR in $0 \sim 20$ dB. For $B = 12$, ZF-MMT provides throughput gain for SNR in $-5 \sim 25$ dB, and the gain is 25% for SNR=$10 \sim 15$ dB.

Remark 11. *The numerical results in this section provide the following insights:*

1. *As shown in Figure 3.6, the effect of channel quantization error dominates the performance of the given system. This can also be inferred from* (3.13)*. Each interference term due to delay/mobility has variance* $\sigma_D^2 = \epsilon_{e,u}^2$*. Each interference term due to channel quantization error has variance* $\sigma_Q^2 = \rho_u^2 2^{-\frac{B}{N_t - 1}}$*. For* $v = 10$ *km/hr,* $f_c = 2.1$ *GHz, and* $\tau = 2$ *msec, we have* $\sigma_D^2 = 0.0295$*, and* $\sigma_Q^2 = 0.3851$ *with* $B = 4$*,* $\sigma_Q^2 = 0.1528$ *with* $B = 8$*, so* $\sigma_D^2 \ll \sigma_Q^2$*. To get* $\sigma_D^2 \approx \sigma_Q^2$*, we need* $B = 15$*. This shows the number of feedback bits in LTE (*$B = 4$*) is not big enough for MU-MIMO, and we need to increase it (*$B > 6$*).*

2. *As shown in Figure 3.5(a), for a given delay and give* B*, the mobility plays a significant role, and MU-MIMO should only be used with low mobility (*$v <$ 30

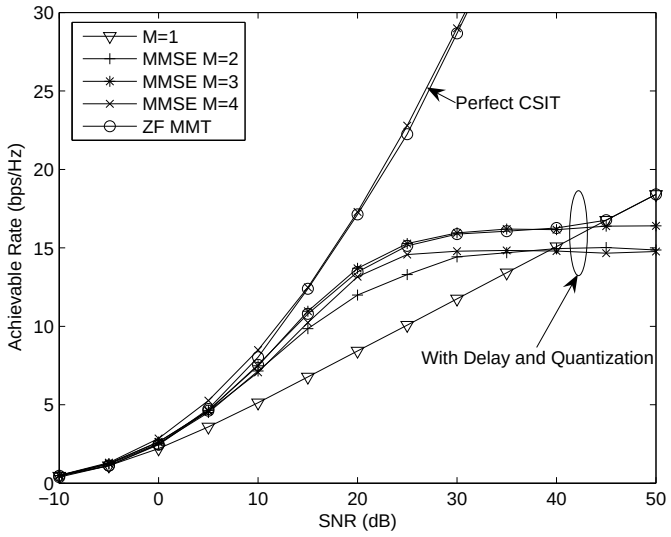


Figure 3.7: Simulation results of MMSE precoding and the ZF precoding with the multi-mode transmission (MMT), $N_t = 4$. For imperfect CSIT, $B = 18$ bits, $f_c = 2$ GHz, $T_s = 1$ msec., and $v = 10$ km/hr.

km/hr).

3. *As delay is related to the radio frame length, it is expected that in the LTE system that has a smaller frame length (1 msec) than the WiMAX system (5 msec) the MU-MIMO is more applicable.*

3.6.2 ZF vs. MMSE Precoding

MMSE precoding, or *regularized ZF precoding*, can increase the throughput at low SNR compared to ZF precoding [113]. In this section, we show that with multi-mode transmission ZF precoding provides performance close to MMSE precoding in the whole SNR range, with perfect or imperfect CSIT.

Figure 3.7 compares the sum rates of ZF precoding employing multi-mode transmission and MMSE precoding. For the perfect CSIT case, the number of active users for MMSE precoding is fixed to be $U = N_t$, as little gain can be achieved by varying the user number. We see that the sum rates of the two systems are very close. This means that multi-mode transmission improves the performance of ZF precoding

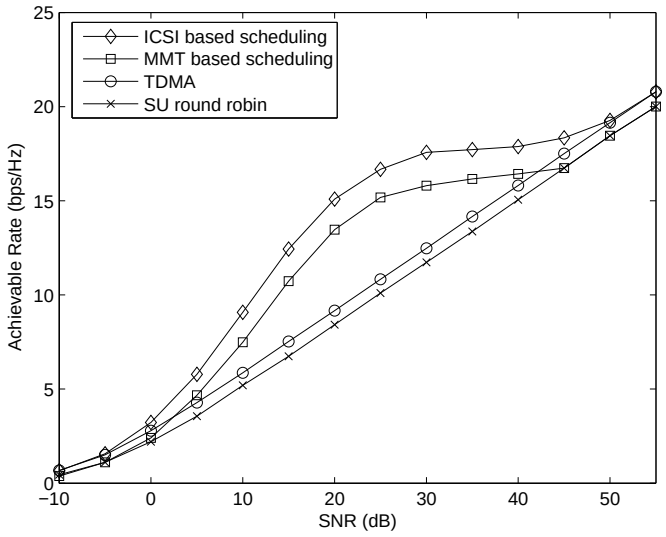


Figure 3.8: Achievable throughputs with different scheduling algorithms in the homogeneous network. The TDMA system schedules the user with the highest instantaneous throughput each time slot. $U = 20$, $B = 18$ bits, $f_c = 2$ GHz, $T_s = 1$ msec., and $v = 10$ km/hr.

and approaches that of MMSE precoding. For imperfect CSIT, simulation results for different modes with MMSE precoding are plotted, showing that mode switching is also required to improve the spectral efficiency for MMSE precoding. If multi-mode transmission is applied for both systems, we see that the performance of ZF precoding approaches that of MMSE precoding.

3.6.3 MMT-based Scheduling vs. ICSI-based Scheduling

In this section, we compare the performance of the proposed MMT-based scheduling and instantaneous CSI (ICSI) based scheduling described in Appendix 3.8.4.

3.6.3.1 Homogeneous Network

The performance comparison is shown in Figure 3.8, with $U = 20$, where it can be seen that the ICSI-based scheduling provides a throughput gain over the

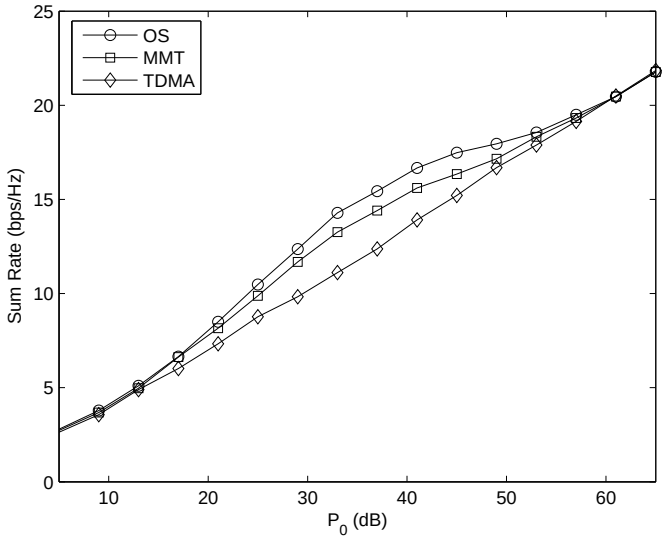
MMT-based one. This is due to the multi-user diversity gain, as the MMT-based scheme does not exploit the time-varying nature of the channel. Nevertheless, the MMT-based scheme provides a significant throughput gain over the SU system at medium SNR.

3.6.3.2 Heterogeneous Network

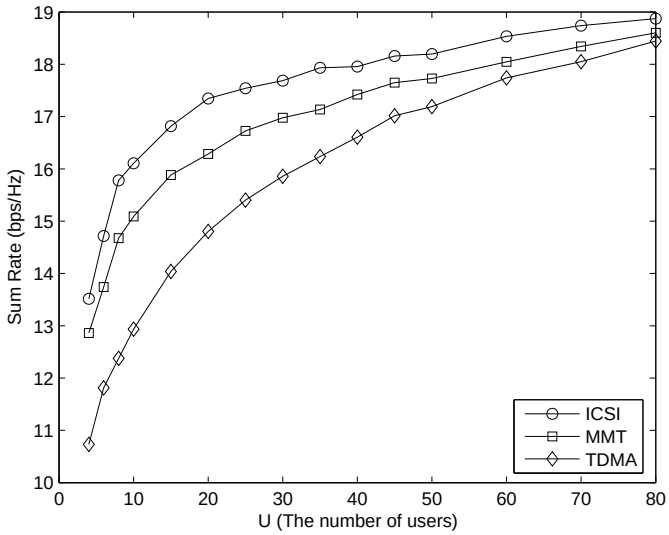
The results for different user scheduling algorithms are shown in Figure 3.9. An isolated cell is considered, where users are uniformly distributed within the cell, so different users have different pathlosses. All the users are assumed to have the same delay and the same number of feedback bits. The path loss model is $P_r(d) = P_0 \left(\frac{d_0}{d}\right)^{\alpha}$, where d is the distance between the BS and the mobile user, d_0 is the reference distance, P_0 is the received power at the reference distance, $P_r(d)$ is the receive power at the mobile, and α is the path loss exponent. In the simulation, the cell radius is 100m, $d_0 = 10$m, and $\alpha = 3.7$.

Figure 3.9(a) compares the sum rates for different P_0 with $U = 30$. We see that both MMT-based and ICSI-based scheduling algorithms provide throughput gain over TDMA for P_0 within -20 dB $\sim$ 20 dB. The MMT-based scheduling provides the sum rate within 1 bps/Hz to the ICSI-based one, as it also provides multi-user diversity gain by selecting users considering different path losses. As shown in [52], path loss determines the multi-user diversity gain in the heterogeneous network. Note that the MMT-based algorithm only requires maximally N_t users to feed back the instantaneous CSI.

Figure 3.9(b) compares the sum rates for different user numbers with $P_0 = 10$ dB. We see that both MMT-based and ICSI-based scheduling algorithms provide a significant throughput gain over TDMA when U is not large ($\lesssim 50$). The MMT-based scheduling provides the sum rate close to the ICSI-based one, within 1 bps/Hz. As U grows, more and more users will be around the central BS, and the BS will mainly select from these interior users that have high SNRs. As the throughput of the MU mode is limited by the residual inter-user interference at high SNR, the BS will tend



(a) Different SNR, $U = 30$.



(b) Different U, $P_0 = 47$ dB.

Figure 3.9: Comparison of TDMA, MMT-based scheduling, and ICSI-based scheduling in a single-cell scenario, $N_t = 4$. $B = 18$ bits, $f_c = 2$ GHz, $T_s = 1$ msec., and $v = 10$ km/hr. The TDMA scheduling selects the user with the highest instantaneous throughput each time slot.

to select a single user. As shown in Figure 3.9(b), as U increases, the sum rates for ICSI-based and MMT-based scheduling algorithms converge to that of the TDMA system.

3.7 Conclusion

In this chapter, we propose a multi-mode transmission technique that adaptively adjusts the number of active users based on the achievable ergodic rate. Considering transmit array gain, spatial division multiplexing gain and residual interference, multi-mode transmission improves the spectral efficiency over SU-MIMO and the dual-mode switching system in Chapter 2 at medium SNR. It is shown that the full mode $M^\star = N_t$ will normally not be activated as it has the highest residual interference and no array gain. User scheduling algorithms considering multi-mode transmission are proposed, which schedule users based on average SNR, codebook size, and normalized Doppler frequency. The proposed algorithms significantly reduce the feedback amount, and provide throughput close to the one based on instantaneous CSI feedback.

We only consider single-antenna receivers. Future mobile terminals are likely to have multiple receive antennas, so it is interesting to extend the current work to this scenario. It is important to consider the impact of more realistic channel models, such as channel correlation. In addition, we only consider the single-cell scenario. It is of practical importance to investigate the achievable performance of our proposed transmission strategies in practical cellular networks with additional impairments such as other-cell interference.

3.8 Appendix

3.8.1 Useful Results for Rate Analysis

The following lemma provide some useful integrals for the sum rate analysis of the multi-mode transmission system. The proof is omitted due to space limitation.

Lemma 4.

$$\hat{I}_1(a,b,m) = \int_b^{\infty} x^m e^{-ax} dx$$
$$= \begin{cases} e^{-ab}\sum_{i=0}^{m} \frac{m!}{i!}\frac{b^i}{a^{m-i+1}} & m \geq 0 \\ E_1(ab) & m = -1 \\ (-a)^{-m-1}\frac{E_1(ab)}{(-m-1)!} + \frac{e^{-ab}}{b^{-m-1}}\sum_{i=0}^{-m-2}\frac{(-ab)^i(-m-i-2)!}{(-m-1)!} & m \leq -2 \end{cases}, \tag{3.20}$$

$$\hat{I}_2(a,b,m,n) = \int_0^{\infty} \frac{x^m e^{-ax}}{(x+b)^n} dx = e^{ab}\sum_{i=0}^{m}\binom{m}{i}(-b)^{m-i}\hat{I}_1(a,b,i-n), \tag{3.21}$$

$$\hat{I}_3(a,b,m,n) = \int_0^{\infty} \frac{x^m e^{-ax}}{(x+b)^n(x+1)} dx$$
$$= \sum_{i=1}^{n}\frac{(-1)^{i-1}}{(1-b)^i}\cdot\hat{I}_2(a,b,m,n-i+1) + \frac{\hat{I}_2(a,1,m,1)}{(b-1)^n}, \tag{3.22}$$

$$\hat{I}_4(a,b,m,n) = \int_b^{\infty} \frac{x^m}{(x+a)^n} dx = \sum_{i=0}^{m}(-a)^{m-i}\frac{(a+b)^{i-n+1}}{n-i-1}, 0 < m < n, \tag{3.23}$$

$$\hat{I}_5(a,b,m) = \int_b^{\infty} \frac{1}{x(x+a)^m} dx = a^{-m}\ln\left(\frac{a+b}{a}\right) - \sum_{i=1}^{m-1}\frac{a^{-i}(a+b)^{-m+i}}{m-i}, n > m, \tag{3.24}$$

$$\hat{I}_6(a,m,n) = \int_0^{\infty} \frac{x^m}{(x+a)^n(x+1)} dx$$
$$= (-1)^m\hat{I}_5(a-1,1,n) + \sum_{i=1}^{m}(-1)^{m-i}\hat{I}_4(a-1,1,i-1,n). \tag{3.25}$$

where $E_1(x)$ is the exponential-integral function of the first order.

3.8.2 Proof of Theorem 6

For the SINR approximation in (3.13), assuming interference terms are independent, and independent of the signal term, $\gamma_{ZF,u}^{(QD)}$ can be written as

$$\gamma_{ZF,u}^{(QD)} \approx \frac{\alpha z}{1+\delta_1 y_1 + \delta_2 y_2} \triangleq x, \tag{3.26}$$

where $\alpha = \frac{\rho_u^2 P}{M}$, $\delta_1 = \frac{\rho_u^2 P \delta}{M}$, $\delta_2 = \frac{\epsilon_{e,u}^2 P}{M}$, $y_1 \sim \chi_{2L}^2$, $y_2 \sim \chi_{2L}^2$, $z \sim \chi_{2(N_t-L)}^2$, $L = M-1$, and y_1, y_2, z are independent of each other.

Let $y = \delta_1 y_1 + \delta_2 y_2$, then the pdf of y, which is the sum of two independent chi-square random variables, is given as [129]

$$p_Y(y) = e^{-y/\delta_1} \sum_{i=0}^{L-1} a_i^{(1)} y^i + e^{-y/\delta_2} \sum_{i=0}^{L-1} a_i^{(2)} y^i = \sum_{j=1}^{2} \sum_{k=0}^{L-1} a_k^{(j)} y^k e^{-y/\delta_j}, \tag{3.27}$$

where

$$a_i^{(1)} = \frac{1}{\delta_1^{i+1}(L-1)!} \left(\frac{\delta_1}{\delta_1 - \delta_2}\right)^L \frac{(2(L-1)-i)!}{i!(L-1-i)!} \left(\frac{\delta_2}{\delta_2 - \delta_1}\right)^{L-1-i}, \tag{3.28}$$

$$a_i^{(2)} = \frac{1}{\delta_2^{i+1}(L-1)!} \left(\frac{\delta_2}{\delta_2 - \delta_1}\right)^L \frac{(2(L-1)-i)!}{i!(L-1-i)!} \left(\frac{\delta_1}{\delta_1 - \delta_2}\right)^{L-1-i}. \tag{3.29}$$

We first derive the cumulative distribution function (cdf) of the random variable x as follows.

$$\begin{aligned}
F_X(x) &= P\left(\frac{\alpha z}{1+y} \le x\right) = \int_0^\infty F_{Z|Y}\left(\frac{x}{\alpha}(1+y)\right) p_Y(y) dy \\
&= \int_0^\infty \left(1 - e^{-\frac{x}{\alpha}(1+y)} \sum_{i=0}^{N_t-L-1} \frac{\left[\frac{x}{\alpha}(1+y)\right]^i}{i!}\right) p_Y(y) dy \\
&= 1 - e^{-x/\alpha} \int_0^\infty \left\{ e^{-\frac{x}{\alpha}y} \sum_{i=0}^{N_t-L-1} \frac{\left[\frac{x}{\alpha}(1+y)\right]^i}{i!} \cdot \sum_{j=1}^{2} \sum_{k=0}^{L-1} a_k^{(j)} y^k e^{-y/\delta_j} \right\} dy \\
&\overset{(a)}{=} 1 - e^{-x/\alpha} \sum_{i=0}^{N_t-L-1} \sum_{j=1}^{2} \sum_{k=0}^{L-1} \frac{a_k^{(j)} (x/\alpha)^i}{i!} \sum_{l=0}^{i} \binom{i}{l} \int_0^\infty \exp\left[-\left(\frac{x}{\alpha} + \frac{1}{\delta_j}\right) y\right] y^{l+k} dy \\
&\overset{(b)}{=} 1 - \sum_{i=0}^{N_t-L-1} \sum_{j=1}^{2} \sum_{k=0}^{L-1} \sum_{l=0}^{i} \frac{a_k^{(j)} (l+k)!}{l!(i-l)!} \cdot \frac{\alpha^{l+k-i+1} e^{-x/\alpha} x^i}{\left(x + \frac{\alpha}{\delta_j}\right)^{l+k+1}},
\end{aligned} \tag{3.30}$$

where step (a) follows binomial expansion of $(y+1)^i$, and step (b) follows the equality $\int_0^\infty y^M e^{-\alpha y} = M! \alpha^{-(M+1)}$.

The expectation of $\ln(1+x)$ on x is derived as follows.

$$\begin{aligned}
&\mathbb{E}_X\left[\ln(1+X)\right] = \int_0^\infty \ln(1+x) dF_X \\
&\overset{(a)}{=} \int_0^\infty \frac{1-F_X(x)}{x+1} dx \\
&= \int_0^\infty \left[\sum_{i=0}^{N_t-L-1} \sum_{j=1}^{2} \sum_{k=0}^{L-1} \sum_{l=0}^{i} \frac{a_k^{(j)}(l+k)!}{l!(i-l)!} \cdot \frac{\alpha^{l+k-i+1} e^{-x/\alpha} x^i}{\left(x+\frac{\alpha}{\delta_j}\right)^{l+k+1}(x+1)} dx \right] \\
&\overset{(b)}{=} \sum_{i=0}^{N_t-L-1} \sum_{j=1}^{2} \sum_{k=0}^{L-1} \sum_{l=0}^{i} \frac{a_k^{(j)}(l+k)!}{l!(i-l)!} \alpha^{l+k-i+1} I_3\left(\frac{1}{\alpha}, \frac{\alpha}{\delta_j}, i, l+k+1\right),
\end{aligned} \tag{3.31}$$

where step (a) follows integration by parts and step (b) follows the expression of $I_3(\cdot,\cdot,\cdot,\cdot)$ in (3.22). Then the achievable sum rate for the u-th user in mode M can be approximated as

$$R_{QD,u}(M) = \mathbb{E}_\gamma\left[\log_2(1+\gamma_{ZF,u}^{(QD)})\right] \approx \log_2(e) \cdot \mathbb{E}_X\left[\ln(1+X)\right],$$

which gives (3.14).

3.8.3 Proof of Theorem 7

From (3.26), the received SINR for the u-th user at high SNR is approximated by

$$\gamma_{ZF,u}^{(D)} \approx \frac{\hat{\alpha} z}{\hat{\delta}_1 y_1 + \hat{\delta}_2 y_2},$$

where $\hat{\alpha} = \rho_u^2$, $\hat{\delta}_1 = \rho_u^2 \delta$, $\hat{\delta}_2 = \epsilon_{e,u}^2$, $z \sim \chi^2_{2(N_t-L)}$, $y_1 \sim \chi^2_{2L}$, $y_1 \sim \chi^2_{2L}$, $L = M-1$, and z, y_1 and y_2 are independent.

Denote $x = \frac{\hat{\alpha} z}{\hat{\delta}_1 y_1 + \hat{\delta}_2 y_2}$, following the same steps as in Appendix 3.8.2, the cdf of x is derived as

$$F_X(x) = 1 - \sum_{i=0}^{N_t-L-1} \sum_{j=1}^{2} \sum_{k=0}^{L-1} \frac{a_k^{(j)} \hat{\alpha}^{k+1}(k+i)!}{i!} \cdot \frac{x^i}{\left(x+\frac{\hat{\alpha}}{\hat{\delta}_j}\right)^{k+i+1}}. \tag{3.32}$$

Then following the steps in (3.31) we get the result in (3.17).

3.8.4 Scheduling Based on Instantaneous CSI Feedback

In this appendix, we describe the scheduling algorithm based on instantaneous CSI feedback used for comparison in this chapter. This is denoted as *Instantaneous CSI (ICSI) based scheduling*, and it requires each user to feed back its instantaneous CSI in each time slot.

First, the BS estimates the instantaneous throughput for different user subsets, for which mobile users need to feed back both the channel direction information (CDI) and channel quality information (CQI). The quantization of CDI is described in Section 2.3.2, and the feedback of CQI can be done in a similar manner as for the limited feedback system [90, 137, 156]. Second, the user selection algorithm is required to select a subset of users based on the throughput estimation. In this section, we propose a modified greedy search user selection algorithm, which explicitly considers the impact of imperfect CSIT.

3.8.4.1 CQI Feedback for ICSI-based Scheduling

We first investigate which kind of CQI should be fed back for the throughput estimation for ICSI-based scheduling. For simplicity, we assume unquantized CQI feedback.

For the SU mode, the BF vector for the u-th user is $\mathbf{f}_u[n] = \hat{\mathbf{h}}_u[n-1]$. The received SNR with expectation on the channel error vector is

$$\begin{aligned}\mathbb{E}_{\mathbf{e}_u}\left[\mathrm{SNR}_u\right] &= \mathbb{E}_{\mathbf{e}_u}\left[P|\mathbf{h}_u^*[n]\mathbf{f}_u[n]|^2\right] \\ &= P\left(\rho_u^2|\mathbf{h}_u^*[n-1]\mathbf{f}_u[n]|^2 + \epsilon_{e,u}^2\right) \\ &= P\left(\rho_u^2 g_u^2[n-1]\left(\cos^2\theta_u\right) + \epsilon_{e,u}^2\right), \end{aligned} \tag{3.33}$$

where $g_u[n-1] = \|\mathbf{h}[n-1]\|$ is the channel gain and $\cos^2\theta_u = |\tilde{\mathbf{h}}_u^*[n-1]\hat{\mathbf{h}}_u[n-1]|^2$. This will be used as the estimation for the received SNR, from which the throughput can be estimated.

For the MU mode, taking the expectation on the channel error vector leads to

a lower bound on the received SINR for the u-th user as follows

$$\begin{aligned}
\mathbb{E}_{\mathbf{e}_u}\left[\gamma_{ZF,u}^{(QD)}\right] &\overset{(a)}{\geq} \frac{\mathbb{E}_{\mathbf{e}_u}\left(P_u|\mathbf{h}_u^*[n]\mathbf{f}_u[n]|^2\right)}{1+\mathbb{E}_{\mathbf{e}_u}\left(\sum_{u'\neq u}P_{u'}|\mathbf{h}_u^*[n]\mathbf{f}_{u'}[n]|^2\right)} \\
&\overset{(b)}{=} \frac{P_u\left(\rho_u^2|\mathbf{h}_u^*[n-1]\mathbf{f}_u[n]|^2+\epsilon_{e,u}^2\right)}{1+\sum_{u'\neq u}P_{u'}\left(\rho_u^2|\mathbf{h}_u^*[n-1]\mathbf{f}_{u'}[n]|^2+\epsilon_{e,u}^2\right)} \\
&> \frac{P_u\rho_u^2|\mathbf{h}_u^*[n-1]\mathbf{f}_u[n]|^2}{\left(1+\sum_{u'\neq u}P_{u'}\epsilon_{e,u}^2\right)+\sum_{u'\neq u}P_{u'}\rho_u^2|\mathbf{h}_u^*[n-1]\mathbf{f}_{u'}[n]|^2} \\
&= \frac{\hat{P}_u|\mathbf{h}_u^*[n-1]\mathbf{f}_u[n]|^2}{1+\sum_{u'\neq u}\hat{P}_{u'}|\mathbf{h}_u^*[n-1]\mathbf{f}_{u'}[n]|^2},
\end{aligned} \tag{3.34}$$

where step (a) follows Jensen's inequality, step (b) follows $\mathbf{h}_u[n]=\rho_u\mathbf{h}_u[n-1]+\mathbf{e}_u[n]$ and the independence between $\mathbf{h}_u[n-1]$ and $\mathbf{e}_u[n]$. To the u-th user, the effective power for each user is $\hat{P}_i=\frac{P_i\rho_u^2}{1+\sum_{u'\neq u}P_{u'}\epsilon_{e,u}^2}$. This lower bound will be used to estimate the instantaneous throughput.

As shown in [75], the interference term in (3.34) can be written as

$$|\mathbf{h}_u^*[n-1]\mathbf{f}_{u'}[n]|^2 = g_u^2[n-1]\left(\sin^2\theta_u\right)\beta(1,N_t-2),$$

where $\sin^2\theta_u$ is the quantization error of the u-th user, $\theta_u=\angle\left(\tilde{\mathbf{h}}_u[n],\hat{\mathbf{h}}_u[n]\right)$, and $\beta(1,N_t-2)$ is a beta $(1,M-2)$ random variable independent of θ_u. Taking the expectation with respect to the beta random variable leads to the following approximation

$$\begin{aligned}
\gamma_{ZF,u}^{(QD)}(M) &\approx \frac{\hat{P}_u|\mathbf{h}_u^*[n-1]\mathbf{f}_u[n]|^2}{1+\sum_{u'\neq u}\hat{P}_{u'}g_u^2[n-1]\left(\sin^2\theta_u\right)\frac{1}{N_t-1}} \\
&\overset{(a)}{\approx} \frac{\hat{P}_u g_u^2[n-1]|\hat{\mathbf{h}}_u^*[n-1]\mathbf{f}_u[n]|^2}{1+\sum_{u'\neq u}\hat{P}_{u'}g_u^2[n-1]\left(\sin^2\theta_u\right)\frac{1}{N_t-1}},
\end{aligned} \tag{3.35}$$

where step (a) approximates the exact channel direction by the quantized version, which is reasonable for small quantization error.

Therefore, from (3.33) and (3.35), each user needs to feed back two scalars, $g_u[n]$ and $\sin^2\theta_u$, from which the BS can estimate the instantaneous throughput for both SU and MU modes.

3.8.4.2 Modified Greedy User Selection Algorithm

The conventional greedy search algorithm was proposed for the system with perfect CSIT [42], which is low-complexity and provides near-optimal performance with perfect CSIT. Starting from the user with the largest channel gain, in each following step, one user is selected from the unselected user set which adds the maximum performance gain, and the process stops when no more user can be added or the performance metric begins to decrease. However, we find that this algorithm does not work well with imperfect CSIT, as it stops searching at the lowest MU mode $M = 2$ which usually has a lower rate than the SU mode. In fact, higher modes may provide a higher throughput than the SU mode, which can be verified through simulation. Based on this fact, we modify the greedy search algorithm to address this problem: instead of stopping at some intermediate mode, it searches over all the mode types, from the SU mode $M = 1$ to the full MU mode $M = N_t$. It is summarized as follows.

1. Initially, set $\mathcal{S}_i = \emptyset$, $i = 0, 1, \cdots, N_t$, and $R_i = 0$, $i = 1, 2, \cdots, N_t$. Set $\mathcal{U} = \{1, 2, \cdots, U\}$.

2. For $i = 1 : N_t$
   Set $R_{old} = 0$.

   (a) for every $u \in \mathcal{U}$

      i. $\hat{\mathcal{S}} = \mathcal{S}_{i-1} + \{u\}$.

      ii. Estimate the sum rate of the user set $\hat{\mathcal{S}}$, denoted as R_{new}.

      iii. if $R_{new} > R_{old}$, set $R_{old} = R_{new}$, and $\hat{u} = u$.

   (b) Let $\mathcal{S}_i = \mathcal{S}_{i-1} + \{\hat{u}\}$, $\mathcal{U} = \mathcal{U} - \{\hat{u}\}$, and $R_i = R_{old}$.

3. $I = \arg\max_{i=1,2,\cdots,N_t} R_i$. $\mathcal{S}_I$ is the scheduled user set.

Chapter 4

Networked MIMO with Clustered Linear Precoding

The performance of conventional cellular MIMO networks is greatly degraded by other-cell interference. From information theoretic perspective, if it is possible to coordinate across all the base stations (BSs), the uplink becomes a multi-access channel and the downlink becomes a broadcast channel, as opposed to an interference channel. This technique is called *BS coordination* or *multi-cell processing*. However, it is unrealistic to perform BS coordination on a global scale due to excessive data information exchange, global CSI and processing complexity requirements. For example, a typical urban cellular network might have 500–1000 BSs. Clearly, coordination must be restricted to a small, neighboring cluster of BSs. Furthermore, the gain from joint transmission with a distant BS is negligible. Therefore, in this chapter we propose a cluster based BS coordination strategy, which includes full intra-cluster coordination and limited inter-cluster coordination.

4.1 Related Work

Intercell scheduling, where neighboring BSs cooperatively schedule their transmissions, is a practical coordination strategy to reduce interference, as each time slot only one BS in each cluster is transmitting and it only requires message change comparable to that for handoff. In [33], it was shown that one major advantage of intercell scheduling compared with conventional frequency reuse is the expanded multiuser diversity gain. However, the interference reduction is at the expense of a transmission duty cycle, and it does not make full use of the available spatial degrees of freedom. Essentially, intercell scheduling is a time-sharing scheme, and as shown in [77, 125], for

a block Rayleigh fading channel the asymptotic gain of the capacity-achieving dirty paper coding (DPC) scheme [36] over time-sharing for the sum rate is $\frac{N_t}{\min(N_t, N_r)}$, where N_t and N_r are the numbers of transmit and receive antennas, respectively. As the total number of transmit antennas at BSs in a cluster may be much higher than the number of receive antennas at each mobile, MU-MIMO is more attractive than time-sharing.

Recently, *BS coordination* has been proposed as an effective technique to mitigate interference in the downlink of multi-cell networks [122]. By sharing information across BSs and designing downlink signals cooperatively, signals from other cells may be used to assist the transmission instead of acting as interference, and the available degrees of freedom are fully utilized. In [123], BS coordination with DPC was first proposed with single-antenna transmitters and receivers in each cell. BS coordination in a downlink multi-cell MIMO network was studied in [159], with a per-BS power constraint and various joint transmission schemes. The maximum achievable common rate in a coordinated network, with zero-forcing (ZF) and DPC, was studied in [47, 84], which demonstrated a significant gain over the conventional single BS transmission. With simplified network models, analytical results were derived for multi-cell ZF beamforming in [131] and for various coordination strategies with grouped cell interior and edge users in [79]. Studies considering practical issues such as limited-capacity backhaul and asynchronous interference can be found in [101, 102, 118, 160].

With BSs coordinating for transmission, it forms an effective MU-MIMO broadcast channel, for which DPC has been shown to be an optimal precoding technique [25, 145, 146, 151, 157]. DPC, while theoretically optimal, is an information theoretic concept that is difficult to implement in practice. A more practical precoding technique for broadcast MIMO channels is *block diagonalization (BD)* [30, 32, 110, 128, 133], which provides each user an interference-free channel with properly designed linear precoding matrices. In addition, it was shown in [127] that BD can achieve a significant part of the ergodic sum capacity of DPC. Therefore, we will apply BD in the multi-cell scenario as the precoding technique for the proposed BS coordination.

Most previous studies on BS coordination assume a global coordination which eliminates inter-cell interference completely. However, in realistic cellular systems, issues such as the complexity of joint processing across all the BSs, the difficulty in acquiring full CSI from all the mobiles at each BS, and time and phase synchronization requirements will make full coordination extremely difficult, especially for a large network. Therefore, it is of great interest to develop coordination schemes at a local scale, to lower the system complexity and maintain the benefits of BS coordination. For the uplink, an overlapping coordination cluster structure was proposed in [143], where each BS is at the center of a unique cluster and coordinated combining is performed to suppress interference for the central BS of each cluster. With such an overlapping cluster structure, each user is in the interior of a cluster and enjoys interference reduction, but the cluster number is as large as the number of BSs and it cannot be easily extended to the downlink. In [18], the downlink coordination over a 3-cell cluster was investigated with both ZF and DPC, but no inter-cluster coordination was considered.

4.2 Contributions

In this chapter, we propose a clustered BS coordination strategy for the downlink of a large cellular MIMO network. With full coordination within the same cluster, the available spatial degrees of freedom[1] are greatly increased, which are then used to reduce inter-cluster interference and exploit the sum rate gain. This strategy consists of a full intra-cluster coordination and a limited inter-cluster coordination. The intra-cluster coordination results in precoding across BSs within the same cluster for MU-MIMO, while the inter-cluster coordination is used to pre-cancel interference for the users at the edge of neighboring clusters. In this way, interferences for both

[1]In this chapter, the definition of the number of spatial degrees of freedom follows [138]. It represents the dimension of the transmitted signal as modified by the MIMO channel, and is equal to the rank of the channel matrix when it has full rank. Therefore, for a point-to-point link with N_t transmit antennas and N_r receive antennas, it is $\min(N_t, N_r)$; for multiuser MIMO channels with K users, it is $\min(N_t, KN_r)$; for BS coordination system with B BSs, it is $\min(BN_t, KN_r)$.

cluster interior and cluster edge users are efficiently mitigated. Meanwhile, the system complexity and CSI requirements at the BSs, which are on a cluster scale, are greatly reduced compared to global coordination. As the main complexity is at the BSs, mobile users can enjoy a simple conventional receiver. In addition, the universal frequency reuse is applied, and there is no need for cell planning.

We apply multi-cell BD as the precoding technique for such coordination. The precoder matrix design is modified from conventional single-cell BD, for which we consider other-cluster interference suppression. In contrast to the classical MIMO broadcast channel, the BS coordination system has a per-BS power constraint. As there is no closed-form solution for the power allocation problem with such a power constraint, three different power allocation algorithms are proposed. For inter-cluster coordination, we show that there is a tradeoff between fairness and sum rate while choosing the inter-cluster coordination area. It is shown that a small cluster size (about 7 cells) can achieve a significant part of the sum rate gain provided by the clustered coordination while greatly reducing channel information feedback compared to global coordination. Simulations show that the proposed coordination strategy improves the sum rate over conventional systems and reduces the impact of interference for cluster-edge users.

The BS coordination considering two classes of users (edge and interior) was also investigated in [79], which derived information-theoretic results based on a simplified Wyner-type circular network model. In this chapter, we consider a more practical setting–a large tesselated 2-D network. We propose clustered coordination based on low-complexity linear precoding, design parameters for such coordination and demonstrate the achievable performance with simulation. We have made some idealized assumptions in this chapter, such as perfect information about channel state and interference. We demonstrate through simulation that the coordination system is sensitive to imperfect channel knowledge.

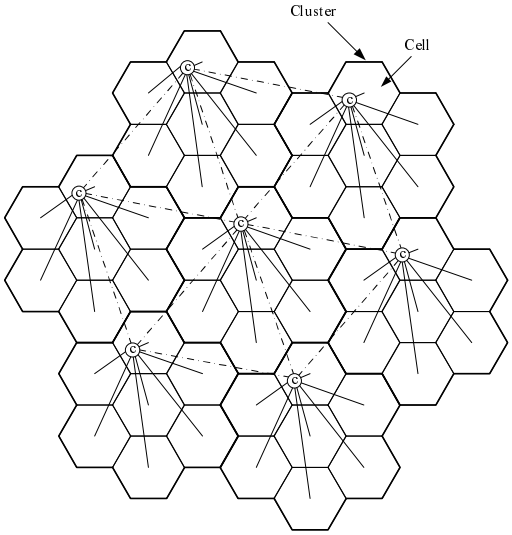


Figure 4.1: An example of the clustered network, with $B = 7$. Node "C" in each cluster is the virtual controller, which means full coordination within each cluster. The dashed line between controllers in neighboring clusters denotes the limited coordination between these clusters.

4.3 System Model

4.3.1 Clustered MIMO Network Structure

Consider a cellular MIMO network, where both BSs and mobile users have multiple antennas, N_t and N_r, respectively. The system parameters used in this chapter are summarized in Table 4.1. We consider a large network, i.e. the number of cells in the network is very large, so it is impractical to do coordination across all the BSs. We propose to divide the network into a number of disjoint clusters, where each cluster contains a group of adjacent cells, as in Figure 4.1. With coordination among the BSs within the same cluster, we effectively increase the number of spatial degrees of freedom, which will be used to suppress interference, including inter-user and inter-cluster interference, and provide sum rate gain.

We apply universal frequency reuse, so the users at the cluster edge may suffer a high degree of interference from neighboring clusters. To efficiently accommodate all the users, we group them into two classes: cluster interior users and cluster edge users. A discussion about user grouping will be given in Section 4.5. To do the

Table 4.1: System Parameters

Symbol	Description
P	the maximum transmit power at each BS
B	number of BSs in each cluster
C	number of clusters we consider
K	number of users per cluster
l_k	length of data symbol for user k
N_t	number of transmit antennas at each BS
N_r	number of receive antennas at each mobile
R	radius of each cell
D_c	coordination distance

proposed clustered coordination, we make several assumptions.

Assumption 1. The BSs within a cluster have perfect CSI of all the users in this cluster, and perfect CSI of the edge users in the neighboring clusters.

For a time-division duplexing (TDD) system, the BS can obtain the downlink CSI through direct uplink channel estimation due to channel reciprocity. For a frequency-division duplexing (FDD) system, the downlink CSI can be obtained by feedback from mobile users, and limited feedback for MU-MIMO is an ongoing topic [68, 75, 114], which we will not explore in this study and perfect CSI is assumed. The assumption of the availability of CSI of the edge users is based on the fact that for handoff such users have CSI of multiple neighboring clusters and can feed back such information. The full CSI of the users in the same cluster is for MU-MIMO precoding to cancel the inter-user interference. The CSI of the edge users in the neighboring clusters is for pre-canceling the inter-cluster interference for these users.

Assumption 2. The BSs within the same cluster can fully share CSI and user data. The BSs in different clusters can exchange traffic information, such as the number of active users and user locations.

The capability of full coordination of the BSs within the same cluster enables doing MU-MIMO precoding across all the BS antennas in this cluster. The limited

coordination between BSs in different clusters can be used for scheduling, e.g. the cluster with a large number of active users may not do inter-cluster coordination for the neighboring clusters.

Assumption 3. BSs within the same cluster are perfectly synchronized in time and phase, and different propagation delays from these BSs to mobile users in this cluster are compensated.

This assumption is to ensure synchronous reception from the home BSs at mobile users. It is difficult to realize perfect synchronization in practice, and the investigation of the impact of asynchronous reception is out of the scope of this dissertation. Recently, there has been some study on this subject [160].

From these assumptions, the system requirements for clustered coordination are based on a cluster scale, which is much lower than that for global coordination, especially in a large network.

4.3.2 Coordination Strategy

Based on the clustered structure and assumptions in the last section, we propose a clustered coordination strategy, including full intra-cluster coordination and limited inter-cluster coordination. The transmission strategies for different user groups are described as follows.

Cluster interior users BSs within the same cluster work together as a "super BS" to serve the interior users in that cluster with MU-MIMO precoding. In this way, there will be no intra-cluster interference, i.e. inter-user interference, for these users. In addition, the interior users are protected to a large degree from inter-cluster interference by path loss.

Cluster edge users Multiple neighboring clusters have channel information of edge users, and they coordinate for the data transmission: one of these clusters is selected

to act as the home cluster to transmit data to such a user, and other neighboring clusters will take this user into consideration when designing precoding matrices. With pre-cancelation of intra-cluster interference provided by the home cluster and pre-cancelation of inter-cluster interference at other neighboring clusters, there will be no interference for this edge user from those clusters.

With such a coordination strategy, the interference for both cluster interior and cluster edge users are efficiently mitigated. Fractional frequency reuse (FFR) is another technique for interference management where BSs cooperatively schedule users in different downlink bandwidths [4–6]. However, FFR is a frequency-domain interference management technique. The proposed coordination strategy is a spatial domain technology that can be implemented with a universal frequency reuse. For a highly-loaded system, FFR alone cannot accommodate all the edge users. Networked MIMO offers another opportunity to serve them.

4.3.3 Received Signal Model

Without loss of generality, we consider the cluster c. The $N_r \times 1$ received signal vector at the kth user in the cluster c is given as

$$\mathbf{y}_k^{(c)} = \underbrace{\sum_{b=1}^{B} \mathbf{H}_k^{(c,b)} \mathbf{T}_k^{(c,b)} \mathbf{x}_k^{(c)}}_{\text{desired signal}} + \underbrace{\sum_{b=1}^{B} \mathbf{H}_k^{(c,b)} \sum_{i=1, i\neq k}^{K} \mathbf{T}_i^{(c,b)} \mathbf{x}_i^{(c)}}_{\text{intra-cluster interference}} + \underbrace{\sum_{\hat{c}=1, \hat{c}\neq c}^{C} \sum_{\hat{b}=1}^{B} \mathbf{H}_k^{(\hat{c},\hat{b})} \sum_{j=1}^{K^{(\hat{c})}} \mathbf{T}_j^{\hat{c},\hat{b}} \mathbf{x}_j^{(\hat{c})}}_{\text{inter-cluster interference}} + \mathbf{n}_k^{(c)} \tag{4.1}$$

where

- $\mathbf{x}_k^{(c)}$ is the $l_k \times 1$ transmitted vector for user k in cluster c. Denote $\bar{\mathbf{x}}^{(c)} = [\mathbf{x}_1^{(c)*} \quad \mathbf{x}_2^{(c)*} \quad \cdots \quad \mathbf{x}_K^{(c)*}]^*$, where $*$ denotes the conjugate transpose of a matrix. The covariance matrix for $\bar{\mathbf{x}}^{(c)}$ is denoted as $\mathbf{Q}^{(c)} = \mathbb{E}[\bar{\mathbf{x}}^{(c)} \bar{\mathbf{x}}^{(c)*}]$.

- $\mathbf{H}_k^{(c,b)}$ is the $N_r \times N_t$ channel matrix from BS b in cluster c to user k.

- $\mathbf{T}_k^{(c,b)}$ is the $N_t \times l_k$ precoding matrix for user k at the bth BS in cluster c.
- $\mathbf{n}_k^{(c)}$ is the additive white Gaussian nose at user k in cluster c, with zero mean and variance $\mathbb{E}(\mathbf{n}_k^{(c)}\mathbf{n}_k^{(c)*}) = \sigma_n^2\mathbf{I}_{N_r}$.

Because the B BSs within this cluster coordinate to work as a super BS, the signal model can be written as

$$\mathbf{y}_k^{(c)} = \mathbf{H}_k^{(c)} \sum_{i=1}^{K} \mathbf{T}_i^{(c)}\mathbf{x}_i^{(c)} + \mathbf{n}_k^{(c)} + \sum_{\hat{c}=1,\hat{c}\neq c}^{C} \mathbf{H}_k^{(\hat{c})} \sum_{j=1}^{K^{(\hat{c})}} \mathbf{T}_j^{(\hat{c})}\mathbf{x}_j^{(\hat{c})} \tag{4.2}$$

where $\mathbf{H}_k^{(c)} = [\mathbf{H}_k^{(c,1)}, \mathbf{H}_k^{(c,2)}, \cdots, \mathbf{H}_k^{(c,B)}]$ is the $N_r \times N_t B$ aggregate channel transfer matrix from the super BS to user k, and

$$\mathbf{T}_k^{(c)} = [\mathbf{T}_k^{(c,1)*}, \mathbf{T}_k^{(c,2)*}, \cdots, \mathbf{T}_k^{(c,B)*}]^*$$

is the aggregate transmit precoder for user k over all B BSs. Unlike traditional downlink with co-located MIMO channels, the channel gains from any two antennas at different BSs are guaranteed to be independent.

Denote $\mathbf{z}_k^{(c)} = \mathbf{n}_k^{(c)} + \sum_{\hat{c}=1,\hat{c}\neq c}^{C} \mathbf{H}_k^{(\hat{c})} \sum_{j=1}^{K^{(\hat{c})}} \mathbf{T}_j^{(\hat{c})}\mathbf{x}_j^{(\hat{c})}$ as the sum of the noise and interference from other clusters, the covariance matrix of which is

$$\begin{aligned}\mathbf{R}_k^{(c)} &= \sigma_n^2\mathbf{I}_{N_r} + \sum_{\hat{c}=1,\hat{c}\neq c}^{C} \sum_{j=1}^{K^{(\hat{c})}} \mathbf{H}_k^{(\hat{c})}\mathbf{T}_j^{(\hat{c})}\mathbb{E}[\mathbf{x}_j^{(\hat{c})}\mathbf{x}_j^{(\hat{c})*}]\mathbf{T}_j^{(\hat{c})*}\mathbf{H}_k^{(\hat{c})*} \\ &= \sigma_n^2\mathbf{I}_{N_r} + \sum_{\hat{c}=1,\hat{c}\neq c}^{C} \sum_{j=1}^{K^{(\hat{c})}} \mathbf{H}_k^{(\hat{c})}\mathbf{T}_j^{(\hat{c})}\mathbf{Q}_j^{(\hat{c})}\mathbf{T}_j^{(\hat{c})*}\mathbf{H}_k^{(\hat{c})*}. \end{aligned} \tag{4.3}$$

Assumption 4. The interference plus noise covariance matrix is perfectly known at the mobile users and BSs in the same cluster.

This covariance matrix can be estimated at mobile users by various methods, including the usage of silent period of the desired signal [109], the usage of pilot signal [82] and blind estimation [63] according to multiple access strategies. After such estimation, each user will feed back it to the BS, which will be used to design precoding matrix.

4.4 Clustered Multi-cell BD

In the proposed coordination strategy, both cluster interior and cluster edge users are served by multi-cell BD with pre-cancelation at the "super BS". BD is a linear precoding technique for downlink MU-MIMO systems, and single-cell BD has been well studied [30, 32, 110, 128, 133]. A major difference between multi-cell BD and single-cell BD is the power constraint. While single-cell BD has a total power constraint (TPC), each BS in the cluster has its own power constraint, so multi-cell BD has a per-BS power constraint (PBPC). In this section, we will design the clustered multi-cell BD, which can be separated into two parts: the precoding matrix design and the power allocation design. The design of the precoding matrix will consider other-cell interference (OCI) and follow the algorithm proposed in [128], which combines interference whitening at the receiver and a statistical OCI-aware precoder at the transmitter to reduce OCI and is shown to provide better sum rate performance than conventional BD. For the power allocation, three different algorithms will be proposed for PBPC.

4.4.1 Precoding Matrix Design

To suppress other-cell interference, we apply an $N_r \times N_r$ whitening filter $\mathbf{W}_k^{(c)}$ at the receiver for each user, which is shown to be related with $\mathbf{R}_k^{(c)}$ as [128]

$$\left[\mathbf{R}_k^{(c)}\right]^{-1} = \mathbf{W}_k^{(c)}\mathbf{W}_k^{(c)*}.$$

With this whitening filter, the received signal for user k after post-processing is

$$\mathbf{r}_k^{(c)} = \mathbf{W}_k^{(c)}\mathbf{H}_k^{(c)}\sum_{i=1}^{K}\mathbf{T}_i^{(c)}\mathbf{x}_i^{(c)} + \mathbf{W}_k^{(c)}\mathbf{z}_k^{(c)} = \hat{\mathbf{H}}_k^{(c)}\sum_{i=1}^{K}\mathbf{T}_i^{(c)}\mathbf{x}_i^{(c)} + \hat{\mathbf{z}}_k^{(c)} \tag{4.4}$$

where $\hat{\mathbf{H}}_k^{(c)} = \mathbf{W}_k^{(c)}\mathbf{H}_k^{(c)}$ and $\hat{\mathbf{z}}_k^{(c)} = \mathbf{W}_k^{(c)}\mathbf{z}_k^{(c)}$ are equivalent channel matrix and noise vector.

Based on the equivalent signal model in (4.4), we can get the precoder for multi-cell BD. First, construct the aggregate interference matrix for user k in cluster c as

$$\tilde{\mathbf{H}}_k^{(c)} = \left[\hat{\mathbf{H}}_1^{(c)*} \quad \cdots \quad \hat{\mathbf{H}}_{k-1}^{(c)*} \quad \hat{\mathbf{H}}_{k+1}^{(c)*} \quad \cdots \quad \hat{\mathbf{H}}_K^{(c)*}\right]^*. \tag{4.5}$$

The principle idea of BD is to find the precoding matrix $\mathbf{T}_k^{(c)}$ such that $\tilde{\mathbf{H}}_k^{(c)}\mathbf{T}_k^{(c)} = \mathbf{0}$, which means there is no inter-user interference. Thus $\mathbf{T}_k^{(c)}$ lies in the null space of $\tilde{\mathbf{H}}_k^{(c)}$. A sufficient condition for the existence of a nonzero effective channel matrix for user k, $\hat{\mathbf{H}}_k^{(c)}\mathbf{T}_k^{(c)}$, is that at least one row of $\hat{\mathbf{H}}_k^{(c)}$ is linearly independent of the rows of $\tilde{\mathbf{H}}_k^{(c)}$ [64]. This introduces the constraint that the number of total transmit antennas (BN_t) is no smaller than the number of total receive antennas (KN_r). Therefore, there is a constraint on the total number of users that can be served simultaneously in each cluster [110, 133], specified as follows [2]:

Lemma 5 (User constraint for multi-cell BD)**.** For a clustered MIMO network with B BSs per cluster, the maximum number of users that can be supported simultaneously in each cluster by multi-cell BD is bounded by

$$K_{max} \leq \left\lfloor \frac{BN_t}{N_r} \right\rfloor.$$

where $\lfloor x \rfloor$ is the maximum integer less than or equal to x.

Assuming $K \leq \left\lfloor \frac{BN_t}{N_r} \right\rfloor$, we describe the precoding matrix design as follows. Let $\tilde{l}_k = \text{rank}(\tilde{\mathbf{H}}_k^{(c)})$, and denote the singular value decomposition (SVD) of $\tilde{\mathbf{H}}_k^{(c)}$ as

$$\tilde{\mathbf{H}}_k^{(c)} = \tilde{\mathbf{U}}_k^{(c)}\tilde{\mathbf{\Lambda}}_k^{(c)}[\tilde{\mathbf{V}}_{k,1}^{(c)}\ \tilde{\mathbf{V}}_{k,0}^{(c)}]^*,$$

where $\tilde{\mathbf{V}}_{k,1}^{(c)}$ contains the first $\tilde{l}_k$ right singular vectors and $\tilde{\mathbf{V}}_{k,0}^{(c)}$ contains the last $BN_t - \tilde{l}_k$ right singular vectors. Therefore, $\tilde{\mathbf{V}}_{k,0}^{(c)}$ forms a null space basis of $\tilde{\mathbf{H}}_k^{(c)}$, from which we can get $\mathbf{T}_k^{(c)}$. In this chapter, we assume the number of spatial streams for each user is $l_k = N_r$. If $l_k < N_r$ or there are extra transmit antennas, additional optimization can be done by picking the appropriate precoder subset [31] or doing coordinated beamforming [28].

With the derived $\mathbf{T}_k^{(c)}$, the received signal becomes

$$\mathbf{r}_k^{(c)} = \hat{\mathbf{H}}_k^{(c)}\mathbf{T}_k^{(c)}\mathbf{x}_k^{(c)} + \hat{\mathbf{z}}_k^{(c)}.$$

[2] If antenna selection or a decoding matrix is applied at the mobile user, it is possible to support more users than this bound [28, 110], which we will not consider in this chapter.

Denote $\bar{\mathbf{T}}_b^{(c)} = [\mathbf{T}_1^{(c,b)} \quad \mathbf{T}_2^{(c,b)} \quad \cdots \quad \mathbf{T}_K^{(c,b)}]$ as the submatrix associated with BS b. Then the transmit power constraint for each BS is

$$\mathrm{Tr}(\bar{\mathbf{T}}_b^{(c)}\mathbf{Q}^{(c)}\bar{\mathbf{T}}_b^{(c)*}) \leq P.$$

The achievable sum rate per cell for the clustered multi-cell BD is then given by [128]

$$R_{CBD} = \max_{\mathrm{Tr}(\bar{\mathbf{T}}_b^{(c)}\mathbf{Q}^{(c)}\bar{\mathbf{T}}_b^{(c)*})\leq P} \frac{1}{B}\sum_{k=1}^{K}\log_2\left|\mathbf{I}_{N_r} + \hat{\mathbf{H}}_k^{(c)}\mathbf{T}_k^{(c)}\mathbf{Q}_k^{(c)}\mathbf{T}_k^{(c)*}\hat{\mathbf{H}}_k^{(c)*}\right|, \tag{4.6}$$

where $\mathbf{Q}_k^{(c)}$ is the covariance matrix for $\mathbf{x}_k^{(c)}$, and $\left[\mathbf{Q}_1^{(c)*}, \mathbf{Q}_2^{(c)*}, \cdots; \mathbf{Q}_K^{(c)*}\right]^* = \mathbf{Q}^{(c)}$.

Denote the SVD of the effective channel $\hat{\mathbf{H}}_k^{(c)}\mathbf{T}_k^{(c)}$ as

$$\hat{\mathbf{H}}_k^{(c)}\mathbf{T}_k^{(c)} = \mathbf{U}_k^{(c)}\begin{bmatrix}\mathbf{\Lambda}_k^{(c)} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\mathbf{V}_k^{(c)},$$

where $\mathbf{\Lambda}_k^{(c)} = \mathrm{diag}(\lambda_{k,1}, \cdots, \lambda_{k,r_k})$, and $r_k = \mathrm{rank}(\mathbf{H}_k^{(c)}\mathbf{T}_k^{(c)})$. Let $\mathbf{\Lambda}^{(c)} = \mathrm{blockdiag}\{\mathbf{\Lambda}_1^{(c)}, \cdots, \mathbf{\Lambda}_K^{(c)}\}$. Then the sum rate can be written as

$$R_{CBD} = \max_{\mathrm{Tr}(\bar{\mathbf{T}}_b^{(c)}\mathbf{Q}^{(c)}\bar{\mathbf{T}}_b^{(c)*})\leq P} \frac{1}{B}\log_2 |\, \mathbf{I} + \mathbf{\Lambda}^{(c)}\tilde{\mathbf{Q}}^{(c)}\mathbf{\Lambda}^{(c)*} \,| \tag{4.7}$$

where $\tilde{\mathbf{Q}}^{(c)} = \mathbf{V}^{(c)*}\mathbf{Q}^{(c)}\mathbf{V}^{(c)}$, and $\mathbf{V}^{(c)} = \mathrm{blockdiag}\{\mathbf{V}_1^{(c)}, \cdots, \mathbf{V}_K^{(c)}\}$.

4.4.2 Power Allocation with PBPC

For the power allocation with PBPC, we propose one optimal and two sub-optimal schemes: user scaling and scaled water-filling. Both the optimal scheme and user scaling scheme are convex optimization problems, and the scaled water-filling scheme is modified from the conventional water-filling power allocation algorithm.

4.4.2.1 Optimal Power Allocation

The optimal power allocation matrix to maximize (4.7) is a diagonal matrix [136], denoted as $\tilde{\mathbf{Q}}_{OPT}^{(c)} = \mathrm{diag}(\gamma_{1,1}, \gamma_{1,2}, \cdots, \gamma_{1,l_1}, \gamma_{2,1}, \cdots, \gamma_{K,l_K})$. The corresponding achievable sum rate is given as

$$R_{CBD} = \max_{\mathrm{Tr}(\bar{\mathbf{T}}_b^{(c)}\mathbf{Q}^{(c)}\bar{\mathbf{T}}_b^{(c)*})\leq P} \frac{1}{B}\sum_{k=1}^{K}\sum_{l=1}^{l_k}\log\left(1+\lambda_{k,l}^2\gamma_{k,l}\right).$$

The power constraint can be rewritten as

$$\sum_{k=1}^{K}\sum_{l=1}^{l_k}\|\mathbf{t}_{k,l}^{(c,b)}\|^2\gamma_{k,l}\leq P, b=1,\cdots,B$$

where $\mathbf{t}_{k,l}^{(c,b)}$ is the lth column of $\mathbf{T}_k^{(c,b)}$.

Thus, the optimal power allocation problem with PBPC can be formulated as

$$\begin{aligned} R_{CBD} = &\max_{\gamma_{k,l}} \frac{1}{B}\sum_{k=1}^{K}\sum_{l=1}^{l_k}\log\left(1+\lambda_{k,l}^2\gamma_{k,l}\right) \\ \text{subject to} &\begin{cases} \sum_{k=1}^{K}\sum_{l=1}^{l_k}\|\mathbf{t}_{k,l}^{(c,b)}\|^2\gamma_{k,l}\leq P, b=1,\cdots,B \\ \gamma_{k,l}\geq 0, l=1,\cdots,l_k, k=1,\cdots,K. \end{cases} \end{aligned} \tag{4.8}$$

For this optimization problem, the power constraints for different users are coupled. It makes the design of the multi-cell BD precoder fundamentally different from its single-cell counterpart. Similar problems with per-antenna power constraints were studied in [17, 158], and the BS coordination system with PBPC appeared in [159]. It was shown in [17] that for the special case of $B = 1$ or 2 this problem can be efficiently solved by the waterfilling algorithm or by characterizing the intersections of the hyperplane constraints, which, however, cannot be easily extended to the more general case. To the best of our knowledge, at this point no closed-form solutions such as waterfilling are available for the optimization problem (4.8).

The objective function, however, is concave and the constraint functions are linear, so this is a convex optimization problem and can be solved numerically, e.g. with the interior-point method [19]. Recently, a more efficient algorithm was proposed in [24], where the dual problem of (4.8) is solved by a B-dimensional subgradient iteration. However, with multiple transmit and receive antennas, and probably a large cluster size, it is still quite complex to solve this optimization problem. In addition, during the user scheduling process, the problem (4.8) needs to be solved for each user combination. For the greedy search scheduling algorithm applied in this chapter, we need to roughly search over $\mathcal{O}(KBK_{max})$ user combinations [126], and the problem (4.8) needs to be solved for each combination except the single-user sets. Therefore, it is desirable to get more efficient algorithms, especially closed-form

solutions. For this purpose, we propose two sub-optimal algorithms in the following sections.

4.4.2.2 User Scaling (US)

One sub-optimal power allocation scheme is user scaling, for which we weight the precoding matrix for each user, by choosing

$$\tilde{\mathbf{Q}}_{US}^{(c)} = \text{blockdiag}(\mu_1 \mathbf{I}_{L_1}, \mu_2 \mathbf{I}_{L_2}, \cdots, \mu_K \mathbf{I}_{l_K}),$$

where μ_k is to scale the precoding matrix of the kth user to meet the power constraint.

There are several reasons for doing this. First, with fewer weight terms it reduces the complexity for solving the optimization problem compared with the optimal scheme. Second, for each user an equal power allocation only results in a negligible capacity loss compared to the optimal water-filling, especially at high SINR, and with shadowing the power allocation across users plays a more important role than across streams of each user. Third, user scaling makes it easy to adjust transmit power between different users, for example, to meet a fixed rate constraint.

Denote $\omega_k^{(c,b)} = \|\mathbf{T}_k^{(c,b)}\|_F^2$. The optimization problem for the user scaling scheme is

$$\begin{aligned} R_{US} = & \max_{\gamma_{k,l}} \frac{1}{B} \sum_{k=1}^{K} \sum_{l=1}^{l_k} \log\left(1 + \lambda_{k,l}^2 \mu_k\right) \\ \text{subject to} & \begin{cases} \sum_{k=1}^{K} \omega_k^{(c,b)} \mu_k \leq P, b = 1, \cdots, B \\ \mu_k \geq 0, k = 1, \cdots, K. \end{cases} \end{aligned} \tag{4.9}$$

Again, this is a convex optimization problem.

4.4.2.3 Scaled Water-Filling (SWF)

As it is difficult to get an efficient algorithm to solve (4.8) and (4.9), we propose another sub-optimal scheme based on the water-filling algorithm.

First, consider a multi-cell BD system with TPC, whose sum rate is given by

$$R_{TPC} = \max_{\text{Tr}(\mathbf{T}^{(c)} \tilde{\mathbf{Q}}_{TPC}^{(c)} \mathbf{T}^{(c)*}) \leq BP} \frac{1}{B} \log_2 | \mathbf{I} + \mathbf{\Lambda}^{(c)} \tilde{\mathbf{Q}}_{TPC}^{(c)} \mathbf{\Lambda}^{(c)*} |. \tag{4.10}$$

where $\mathbf{T}^{(c)}$ is the aggregated precoding matrix, i.e. $\mathbf{T}^{(c)} = [\mathbf{T}_1^{(c)}, \mathbf{T}_2^{(c)}, \cdots, \mathbf{T}_K^{(c)}]$. The optimal power loading matrix $\tilde{\mathbf{Q}}_{TPC}^{(c)} = \mathbf{\Sigma}^{(c)}$ is derived by water-filling [133]. To meet PBPC, we scale this matrix and choose $\tilde{\mathbf{Q}}_{SWF}^{(c)} = \mu \tilde{\mathbf{Q}}_{TPC}^{(c)}$. The scaling factor $\mu \in (0, 1)$ is given by

$$\mu = \frac{P}{\max_{b=1,2,\cdots,B} \mathrm{Tr}(\bar{\mathbf{T}}_b^{(c)} \tilde{\mathbf{Q}}_{TPC}^{(c)} \bar{\mathbf{T}}_b^{(c)*})}$$

Therefore, the sum rate per cell is given by

$$R_{SWF} = \frac{1}{B} \log_2 | \mathbf{I} + \mu \mathbf{\Lambda}^{(c)} \mathbf{\Sigma}^{(c)} \mathbf{\Lambda}^{(c)*} |. \tag{4.11}$$

Compared to the algorithm in [24], this sub-optimal algorithm has a closed-form solution, which avoids the B-dimensional subgradient iteration. In addition, it will be shown later that the proposed scaled waterfilling algorithm provides near-optimal performance.

4.4.3 Scheduling Schemes

From *Lemma 5*, there is a constraint on the maximum number of users a multi-cell BD system can support simultaneously. Therefore, with a large number of users in each cluster, it is necessary to schedule transmission for a subset of users, according to some performance criterion. The sum rate optimal scheduling algorithm is to exhaustively search over all the possible user combinations and pick the user set which maximizes the chosen performance metric, which is extremely complicated. We propose to use a sum rate based sub-optimal user selection algorithm inspired by [126], which has low complexity and approaches optimal performance.

Let $\mathcal{U}$ and $\mathcal{S}$ denote the sets of unselected and selected users respectively, and f_k denotes the performance metric for user k. The proposed user selection algorithm is described in Table 4.2. This is a greedy algorithm. In each step, one user is selected from the un-selected user set which adds the maximum performance gain, and the process stops when no more user can be added or the performance metric begins to decrease. We consider two different kinds of scheduling, maximum sum rate (MSR) and proportional fairness (PF), for different scenarios.

Table 4.2: User Selection Algorithm

1. Initially, set $\mathcal{S} = \emptyset$ and $\mathcal{U} = \{1, 2, \cdots, K\}$. Set $C_{old} = 0$.
2. While $|\mathcal{S}| < K$ and $|\mathcal{S}| < \frac{BN_t}{N_r}$
   (a) for every $k \in \mathcal{U}$
      i. $\hat{\mathcal{S}} = \mathcal{S} + \{k\}$.
      ii. Calculate $C_{new} = \sum_{s \in \hat{\mathcal{S}}} f_s$.
      iii. if $C_{new} > C_{old}$, set $C_{old} = C_{new}$, and $\hat{k} = k$.
   (b) Let $\mathcal{S} = \mathcal{S} + \{\hat{k}\}$, $\mathcal{U} = \mathcal{U} - \{\hat{k}\}$.

4.5 Inter-cluster Coordination

With the proposed coordination strategy, BSs within a cluster serve their interior users with multi-cell BD, while the neighboring clusters coordinate with each other to serve edge users. It is possible for multiple BSs to transmit data to an edge user, but for simplicity we consider that each user is served by one cluster. In this section, we will describe inter-cluster coordination in detail, and investigate two important system parameters: coordination distance and cluster size.

4.5.1 Inter-cluster Coordination with Multi-cell BD

The main idea of inter-cluster coordination is to do interference pre-cancelation at all the neighboring clusters for the active edge user, and select one cluster to transmit information data to this user. The precoding technique used for inter-cluster coordination is multi-cell BD, the same as for intra-cluster coordination. Each edge user selects a cluster based on the channel state, denoted as the *home cluster*, and feeds back this decision, while the other neighboring clusters act as *helpers* for the data transmission. The remaining clusters are *interferer clusters*. Different kinds of clusters and inter-cluster transmission are illustrated in Figure 4.2.

For the following discussion and simulation, we focus on a home cluster and assume that when this cluster schedules an edge user, the neighboring clusters of this

edge user will always help. This will happen if there are a small number of users in each cluster so that there are spare degrees of freedom at neighboring clusters. With a large number of users, joint scheduling across clusters is required. While we leave the full investigation of such a scheduling problem to future work, we propose a simple two-step approach: first, each cluster does scheduling within its own cluster, and the scheduled edge users inform the neighboring helper clusters; in the second step, each cluster deals with the requests from edge users in the neighboring clusters, and it selects to help some of these users while drops some scheduled users of its own. After this scheduling process, each cluster designs precoding matrices.

To the home cluster, there is no difference between the edge user and interior users, and the BD precoding matrix is designed as in Section III. For helper clusters, the precoding matrix design will be different. Without loss of generality, we consider the precoding matrix design at the helper cluster c_1 for the edge user k_0, which is served by its home cluster c_0. Denote

$$\bar{\mathbf{H}}_{k_1}^{(c_1)} = [\tilde{\mathbf{H}}_{k_1}^{(c_1)*} \quad \hat{\mathbf{H}}_{k_0}^{(c_1)*}]^*,$$

where $\tilde{\mathbf{H}}_{k_1}^{(c_1)}$ is the aggregate interference matrix of user k_1 to all the other active users [3] in the cluster c_1 as in (4.5), and $\hat{\mathbf{H}}_{k_0}^{(c_1)} = \mathbf{W}_{k_0}^{(c_0)}\mathbf{H}_{k_0}^{(c_1)}$ is the effective channel after whitening filter from the cluster c_1 to the edge user k_0.

To pre-cancel the interference for both the edge user k_0 and other active users in the cluster c_1, the precoding matrix $\mathbf{T}_{k_1}^{(c_1)}$ should satisfy the condition $\bar{\mathbf{H}}_{k_1}^{(c_1)}\mathbf{T}_{k_1}^{(c_1)} = \mathbf{0}$, i.e. it should lie in the null space of $\bar{\mathbf{H}}_{k_1}^{(c_1)}$, which can be designed with SVD of $\bar{\mathbf{H}}_{k_1}^{(c_1)}$ in the same way as in Section 4.4. Similar to *Lemma 5*, there is a constraint on the number of users that can be supported simultaneously in the helper cluster, stated as follows:

Lemma 6 (User constraint for the helper cluster)**.** For a helper cluster with B BSs and k_e edge users to help, the maximum number of users that can be supported

[3] The active users in a cluster are the users currently being served.

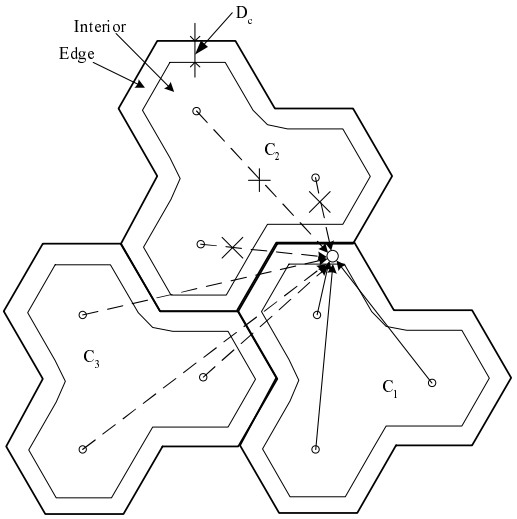


Figure 4.2: An example of inter-cluster coordination, $B = 3$. C_1 is the home cluster, C_2 is the helper cluster, and C_3 is the interferer cluster. Solid lines denote transmissions of information signals and dotted lines are interference, and the cross on the dotted line means that the interference is pre-canceled.

simultaneously by multi-cell BD in this cluster is bounded by

$$K_{max}^{h} \leq \left\lfloor \frac{BN_t}{N_r} \right\rfloor - k_e.$$

Therefore, to serve an edge user with inter-cluster coordination, the total number of users the network can support will be reduced, which induces a tradeoff between mitigating interference for edge users and maximizing the total throughput. This makes the choice of the inter-cluster coordination area important. Actually, the user constraints in *Lemma 5* and *Lemma 6* are due to the constraint on the total spatial degrees of freedom in each cluster, determined by the cluster size and the number of transmit antennas at each BS. To serve an edge user all the neighboring clusters need to provide a certain number of degrees of freedom, which leaves fewer degrees of freedom to serve their own cluster interior users.

4.5.2 Inter-cluster Coordination Distance

In this section, we present one method for grouping the users into cluster interior and cluster edge users, which will be employed in our simulations to illustrate our algorithms' performance. Our proposed metric is based on the channel model, which includes Rayleigh fading, shadowing and path loss, and omnidirectional antennas. With this model, users near the cluster edge will have low signal power and high interference on average, and require inter-cluster coordination. Therefore, we do user grouping based on user locations, and determine an *inter-cluster coordination area* by the *coordination distance*, which is defined as follows and illustrated in Figure 4.2.

Definition 1. Coordination distance, *D_c, is the boundary between interior and edge users. If the distance of the user to the cluster edge is no larger than D_c, this user is classified as a* cluster edge user*; otherwise, it is a* cluster interior user*.*

In a real implementation, this grouping could be performed based on average signal strength measurements (as employed in the handoff algorithm for example). We defer development of measurement based approaches, however, to future work.

Naturally there is a tradeoff when choosing D_c. If D_c is large, more users will be treated as edge users and enjoy a substantial interference reduction, but the total throughout will be reduced as the total number of active users will be reduced. To balance fairness to edge users and the total sum rate, we will investigate the *mean minimum rate* and *effective sum rate*, as a function of D_c.

Mean Minimum Rate Suppose that the mobile users are randomly distributed within each cluster. For a given D_c, for each realization of user locations, denote $R_{min}(D_c)$ as the minimum rate among all the users in the cluster. *Mean minimum rate*[4], $\bar{R}_{min}(D_c)$, is the mean value of $R_{min}(D_c)$, which is mainly determined by the edge users and will increase as D_c increases.

[4]Other similar performance metrics regarding the fairness to the edge users can also be applied, e.g., the achievable rate at a certain outage probability. The results, however, will not change.

Effective sum rate As the edge user is served by multiple neighboring clusters, effectively its rate is shared by those clusters. If there are $N_{c,i}$ clusters serving user i, which is decided by its location and D_c, then the effective rate of this user for each coordinating cluster is $R_i/N_{c,i}$, where R_i is given as follows according to (4.6)

$$R_i = \frac{1}{B}\log_2 \left| \mathbf{I}_{N_r} + \hat{\mathbf{H}}_i^{(c)} \mathbf{T}_i^{(c)} \mathbf{Q}_i^{(c)} \mathbf{T}_i^{(c)*} \hat{\mathbf{H}}_i^{(c)*} \right|. \tag{4.12}$$

The *effective sum rate* for each cluster is defined as

$$R_{sum}(D_c) = \sum_{k=1}^{K} \frac{R_k}{N_{c,k}(D_c)}, \tag{4.13}$$

which will decrease with the increase of D_c as more users become edge users.

For a home cluster, if all the users are interior users, then the effective sum rate is the conventional sum rate for this home cluster; if there is an edge user in this home cluster served by N_c clusters, only $1/N_c$ of this user's rate is counted into the effective sum rate of each serving cluster, including the home cluster. Therefore, the effective sum rate is the same for each cluster in a homogeneous network.

According to these definitions, $\bar{R}_{min}(D_c)$ and $R_{sum}(D_c)$ characterize the opposing objectives of fairness to edge users and total sum throughput. We propose to use a utility function, $U(D_c)$, to evaluate the effect of D_c on both $\bar{R}_{min}$ and R_{sum}.

Definition 2 (Utility Function $U(D_c)$)**.** *The utility function $U(D_c)$ is defined by*

$$U(D_c) = \alpha \frac{\bar{R}_{min}(D_c)}{\max_{D_c} \bar{R}_{min}(D_c)} + (1-\alpha)\frac{R_{sum}(D_c)}{\max_{D_c} R_{sum}(D_c)}, 0 \le \alpha \le 1. \tag{4.14}$$

where α is a variable reflecting the design objective. If it is more valuable to care about edge users, we can pick $\alpha \to 1$; if sum rate is more important, we can pick $\alpha \to 0$. As an example, we pick $\alpha = 1/2$, which means we treat relative changes of $\bar{R}_{min}$ and R_{sum} as of equal value to the system.

Simulation results of $U(D_c)$ for $D_c \in [0, R)$ [5] are shown in Figure 4.3, with $B = 3$, $R = 1$ km and $K = 30$, and interference-free SNR at the cell edge is 18 dB. Totally

[5]When $D_c = R$, the area around the BSs is classified as inter-cluster coordination area, which is not going to be the case as the nearby BS can provide a high SINR for the users in this area. Therefore, we only consider $D_c \in [0, R)$ in this simulation.

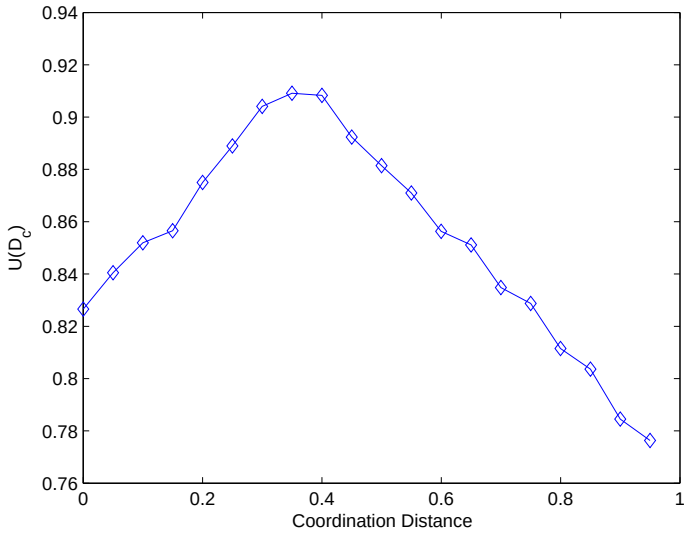


Figure 4.3: $U(D_c)$ for different D_c, $R = 1$ km.

1000 realizations of user locations are run, and for each realization 1000 iterations are simulated with independent channel state. PF scheduling is used to provide fairness, and the scaled water-filling power allocation is used for computational efficiency. From the results we can see that the maximum value is achieved around $D_c = 0.35R$, which will be a proper choice.

Inter-cluster coordination may be designed for criterions other than $\bar{R}_{min}$ and R_{sum}, but the idea of making a good tradeoff between the fairness for the edge users and the sum rate persists.

4.5.3 Cluster Size

With a fixed D_c, if the cluster size is small, the relative coordination area is large and there will be too many cluster edge users which will consume lots of the degrees of freedom and lower the effective sum rate. Alternatively, a large cluster size will have a relatively small coordination area, which has small sum rate loss. However, the requirement of full CSI and synchronization will prohibit a very large cluster size, and due to path loss the users benefit little from those BSs far away. Therefore, to select a suitable cluster size is important for practical systems, which

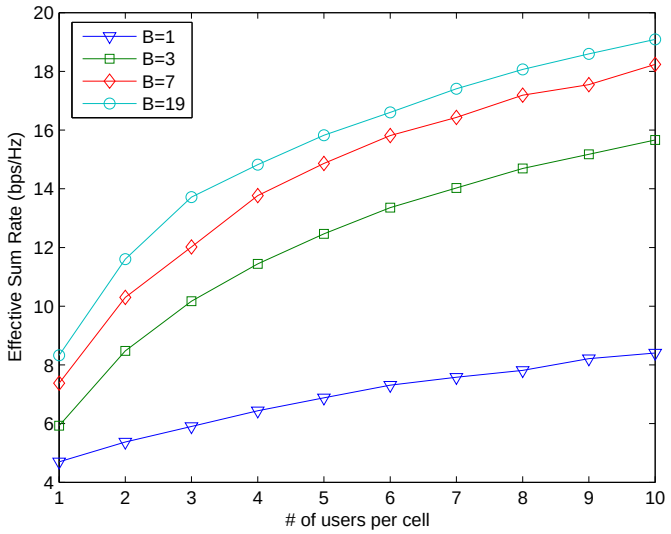


Figure 4.4: Effective sum rate per cell with different cluster size, $B = 1, 3, 7, 19$, $R = 1$ km, and $D_c = 0.35R$. The standard deviation of shadowing is 8dB, the path loss exponent is 3.7.

is also the motivation to propose the clustered coordination.

4.5.3.1 Sum Rates for Different Cluster Sizes

Figure 4.4 shows the effective sum rates per cell for different cluster sizes, $B = 1, 3, 7, 19$, $D_c = 0.35R$, $R = 1$ km, and interference-free SNR at the cell edge is 18 dB. We can see that there is a diminishing gain with the increase of the cluster size: the 3-cell cluster has a much higher sum rate than the 1-cell cluster, and a 7-cell cluster has a rate gain about 2.5 bps/Hz over a 3-cell cluster, while from $B = 7$ to $B = 19$ the sum rate increases about 1 bps/Hz. The lower sum rate for $B = 1$ is due to its relative large edge area. Therefore, a 7-cell cluster can already achieve a significant part of the performance gain of the clustered coordination.

4.5.3.2 CSI Requirement Reduction

The CSI requirement for clustered coordination is on a cluster scale, which is greatly reduced compared to global coordination. With C clusters and B cells

each cluster, totally there are BC BSs in the network. For global coordination, the effective channel matrix for each user is $N_r \times BCN_t$, which is required at each of the BC BSs, while for clustered coordination the dimension of the channel matrix is $N_r \times BN_t$ and it is required at each of the B BSs in that cluster. Therefore, we get the following lemma:

Lemma 7 (CSI Reduction)**.** For a cellular network with C clusters, the amount of CSI required for clustered coordination is $\frac{1}{C^2}$ of that for global coordination.

The amount of required CSI for a 7-cell cluster system is only $\frac{7^2}{19^2}$ of that for a 19-cell cluster, while the performance of the 7-cell cluster system does not degrade much as shown in Figure 4.4, so a cluster size of 7 is a reasonable choice for clustered coordination with the given transmit power.

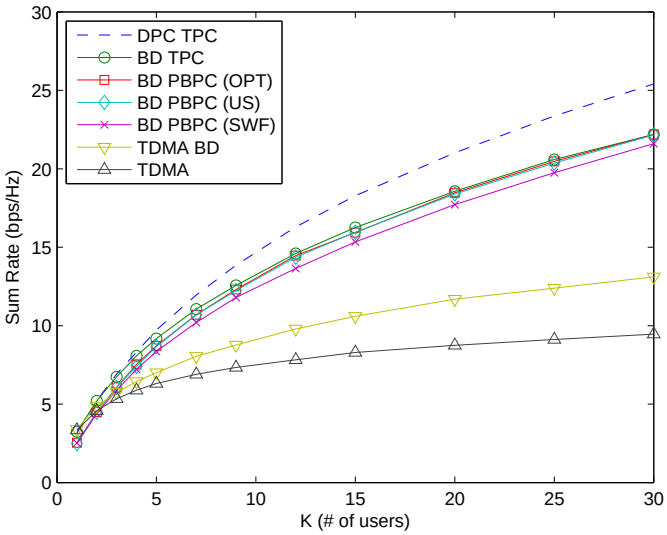
4.6 Numerical Results

In this section, the performance of the proposed coordination strategy is shown via monte carlo simulation. We choose the number of antennas to be $N_t = 4$ and $N_r = 2$. The standard deviation of shadowing is 8 dB, the path loss exponent is 3.7, and the cell radius is 1 km. Other than stated, the interference-free SNR at the cell edge is 18 dB, accounting for path loss and ignoring shadowing and Raleigh fading. We assume all the BSs in other clusters transmit at full power. Mobile users are uniformly distributed within each cluster, and they are associated with clusters based on locations.

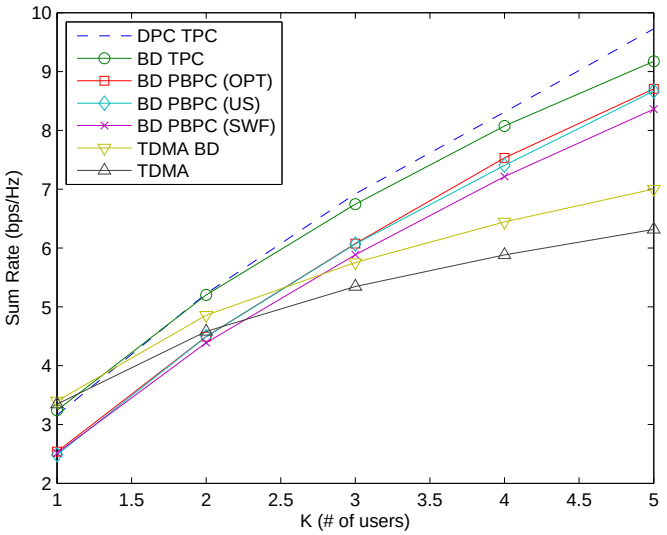
4.6.1 Sum Rates for Different Systems

First, we consider sum rates for different systems with maximum sum rate (MSR) scheduling. Besides the proposed multi-cell BD systems, we also compare with the following systems.

- *Multi-cell DPC with Total Power Constraint (TPC):* This is an upper bound for the downlink channel of multi-cell systems. We assume a total power constraint.



(a) Different K



(b) Small K

Figure 4.5: Sum rate per cell for different systems, with cluster size $B = 3$. "OPT" denotes the optimal power allocation scheme, "US" denotes the user scaling scheme, and "SWF" denotes the scaled water-filling scheme. "DPC TPC" is multi-cell dirty paper coding with total power constraint, and "TDMA" is opportunistic intercell scheduling.

DPC is applied across BSs within the same cluster, and algorithm 2 in [78] is used for power allocation.

- *Multi-cell BD with TPC:* This is similar to the single-cell BD, and the water-filling algorithm can be applied to the aggregated channel for power allocation. This serves as an upper bound for multi-cell BD with PBPC, and can indicate the capacity loss due to the PBPC.
- *TDMA with Intercell Scheduling [33]:* Neighboring BSs cooperatively schedule their transmission, and only one BS is active to serve one user at each time slot.
- *Intercell Scheduling with BD:* Compared to TDMA with intercell scheduling only, this technique allows one BS to serve multiple users at each time slot with BD.

Figure 4.5 compares sum rates for different systems. There are several key observations.

1. The sum rates of multi-cell BD systems are much higher than that of the TDMA system with intercell scheduling, and are pretty close to that of DPC.
2. All multi-cell BD schemes have about the same performance.
3. There is only a marginal rate loss of PBPC to TPC.

4.6.2 Different Power Allocation Schemes

In Figure 4.5, multi-cell BD with different power allocation schemes have close performance, as MSR scheduling selects users in the high SINR regime. If the selected users are not in the high SINR regime, which may occur for other scheduling schemes, the story changes. Figure 4.6 shows results for the scenario where 2 random users are selected in each cell. For this case, user scaling is not as good as the optimal scheme, and BTS scaling has a large rate loss, but the scaled water-filling scheme is still close to the optimal one. Note that we assume full transmit power at BTSs in interfering clusters, which is not the case for scaled water-filling in practical systems.

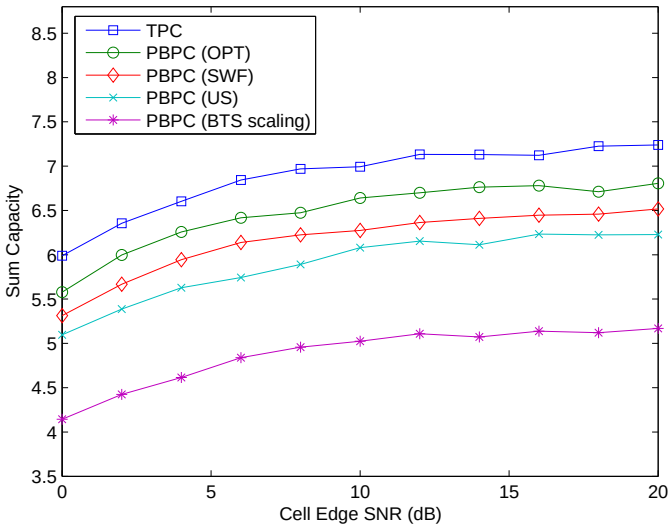


Figure 4.6: Sum rates of multi-cell BD with heterogeneous users, 2 users each cell and $B = 3$.

In addition, modified from conventional water-filling algorithm, scaled water-filling has lower complexity.

4.6.3 Distribution of User Rates

Figure 4.7 shows the cumulative distribution function (CDF) of mean rates for users. There are 30 users uniformly distributed in the cluster, $B = 3$ and $D_c = 0.35R$, and PF scheduling is applied. The simulation setting is similar as that for Figure 4.3. We run 100 realizations for user locations, and for each realization 1000 iterations are simulated with independent channel state and the mean rates are stored. Totally, there are 3000 samples of user rates, with which we can plot the CDF. For example, the rate with 10% outage for intercell scheduling is 0.4 bps/Hz, for intercell scheduling with BD is 0.1 bps/Hz, and for clustered multi-cell BD with and without inter-cluster coordination are 0.6 and 0.8 bps/Hz, respectively. For multi-cell BD with inter-cluster coordination, nearly 60% users have mean rate larger than 1 bps/Hz and 10% users have mean rate larger than 2 bps/Hz, while for intercell scheduling only less than 5% of users can have mean rate larger than 1 bps/Hz.

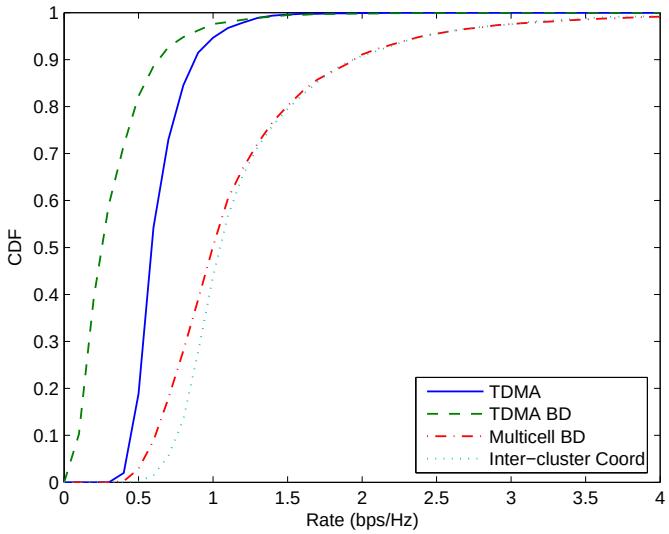


Figure 4.7: CDF of the rates for users in the cluster, $B = 3$, $D_c = 0.35R$.

4.6.4 Imperfect Channel Knowledge

Pilot symbols are required for channel estimation, and such training overhead becomes greater for a larger cluster size. However, there will be inevitable estimation errors, and the simulation results accounting for imperfect channel knowledge are shown in Figure 4.8. The channel estimation model in [30] is used. At the BSs, the available knowledge of the small-scale fading channel matrix of the kth user is given by $\check{\mathbf{H}}_k^{(c,b)} = \mathbf{H}_k^{(c,b)} + \mathbf{E}_k^{(c,b)}$, where $\mathbf{H}_k^{(c,b)}$ is the true channel matrix and $\mathbf{E}_k^{(c,b)}$ is the channel error. Entries of $\mathbf{E}_k^{(c,b)}$ follows i.i.d. complex Gaussian distribution with zero mean and covariance $\sigma_{MSE}^2/2$ per real dimension. The channel knowledge error is denoted as MSE$= 10\log_{10}\sigma_{MSE}^2$ dB. To demonstrate the impact of imperfect CSI, we assume equal MSE for each user. The unequal MSE case is left to future work. We can see that the sum rates for BD systems decrease as MSE increases, while TDMA system with intercell scheduling is not sensitive to channel error, but the sum rates of multicell BD systems are always higher for the simulated range. This is due to the imperfect inter-user interference cancelation with channel error for MU-MIMO systems, and such interference is from the same propagation channel as the information signal, so it will greatly degrade the performance. Therefore, robust

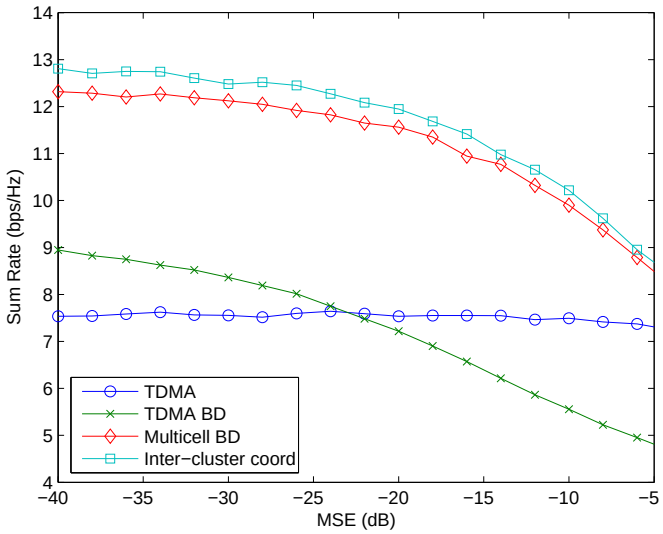


Figure 4.8: Sum rates for different systems with imperfect channel knowledge.

precoding schemes are required in practical systems.

4.7 Conclusion

In this chapter, a clustered BS coordination strategy is proposed to increase the available spatial degrees of freedom for MIMO networks, and thus to reduce interference and increase the sum rate. A cluster structure is formed, and the users are grouped into cluster interior users and cluster edge users, served with different coordination strategies. Cluster interior users are served with intra-cluster coordination, i.e. multi-cell BD, while cluster edge users are served by multiple neighboring clusters to reduce inter-cluster interference. The precoder for multi-cell BD and system parameters for inter-cluster coordination are designed. It is shown that a small cluster size (such as 7) is enough to provide the benefits of the clustered coordination while greatly reducing the amount of channel feedback. Numerical results show that the proposed coordination strategy can provide robust sum rate and edge user rate gains.

There are many practical issues associated with clustered BS coordination,

requiring much future work. Compared with global coordination, the cluster structure reduces the amount of CSI required at the BS, but with multiple antennas at both the BS and mobiles, the amount of CSI is still daunting. Current schemes are sensitive to synchronization and CSI error, which is expected to increase in a cluster system, so robust precoding schemes are needed. In this chapter, we have assumed that all users have perfect knowledge about other-cluster interference. The investigation of the imperfect interference estimation is of practical importance and is a worthy topic of future work. Generally, the analysis of cellular MIMO networks is an open problem, given the randomness of user locations, path loss, and matrix channels with fading and shadowing.

Chapter 5

Adaptive Spatial Intercell Interference Cancellation

The performance of contemporary multicell wireless networks is limited by other-cell interference (OCI), due to cochannel transmission in other cells. This performance degradation is especially severe for users close to the cell edge. *Multicell processing*, or *BS coordination*, is an efficient way to suppress OCI through coordination among multiple BSs. Multicell processing in the downlink can be categorized into two classes:

1. **Coordinated single-cell transmission**: data is transmitted from a single BS, and the OCI suppression is achieved through joint resource allocation among multiple BSs, such as joint power control and user scheduling [51]. Neighboring cells share such information as the offer load in each cell, the channel state information (CSI) of edge users, but no data exchange is required.

2. **Coordinated multicell transmission**: in addition to the information shared in coordinated single-cell transmission, BSs need to exchange user data. A central unit (CU) is normally needed for joint processing of data transmission for BSs that join the coordination, so each user receives data from multiple BSs. Ideally, assuming full CSI and all the data available at the CU, coordinated multicell transmission is able to eliminate all the OCI and the system is no longer interference-limited [122, 132]. The clustered coordination proposed in Chapter 4 is a type of coordinated multicell transmission.

Although coordinated multicell transmission is able to provide a considerable performance gain through efficiently exploiting the available spatial degrees of freedom, it requires a significant amount of inter-BS information exchange and is of high

complexity. This would be quite challenging for practical implementation. First, the large overhead and information exchange would put onerous demands on backhaul capacity; second, precise synchronization among different BSs is required; third, the CSI from each mobile user is required at all the coordinated BSs, which makes CSI estimation and feedback daunting.

On the other hand, coordinated single-cell transmission is of lower overhead and complexity, as no inter-BS data exchange is required, and normally each user needs to provide instantaneous or statistical CSI only to some of its neighboring BSs. In this chapter, we consider a multicell network with multiple antennas at each BS. Coordinated single-cell transmission is applied in the form of intercell interference cancellation (ICIC) through zero-forcing (ZF) precoding. Canceling OCI for neighboring cells consumes available spatial degrees of freedom, so it reduces the received signal power for the home user, and is not necessarily optimal at each BS. We propose an adaptive ICIC strategy where multiple BSs jointly select transmission techniques based on user locations. Each BS only needs to exchange the location of its home user with neighboring BSs, and the CSI of users in neighboring cells is required only when ICIC is applied.

5.1 Related Work

Coordinated multicell transmission, also called *networked MIMO*, has recently drawn significant attention. In a networked MIMO system, multiple coordinated BSs effectively form a "super BS", which transforms an interference channel into a MIMO broadcast channel, with a per-BS power constraint [17, 85, 158]. The optimal dirty paper coding (DPC) [36, 151] and sub-optimal linear precoders have been developed for networked MIMO [48, 73, 83, 123, 159, 166]. With simplified network models, analytical results have appeared in [7, 66, 79, 131].

In practice, the major challenges for networked MIMO concern complexity and overhead. For example, the requirement for CSI grows in proportion to the number of BS antennas, the number of BSs, and the number of users. The complexity of

joint processing also grows with the network size. To limit the complexity and CSI requirements, cluster-based coordination is one approach [18, 111, 143, 166]. To reduce the complexity, distributed decoding and beamforming for networked MIMO systems were proposed in [8, 105, 106]. In [50, 130], BS coordination with hybrid channel knowledge was investigated, where each BS has full information of its own CSI and statistical information of other BSs' channels. Limited backhaul capacity [101, 118] and synchronization [81, 160] have also been treated to some extent. A WiMAX-based implementation of networked MIMO was done in [144], for both uplink and downlink in the indoor environment.

Coordinated single-cell transmission, where the traffic data for each user comes from a single BS, is of lower complexity, requires less inter-BS information exchange, and has lower CSI requirements. Intercell scheduling has been shown to be able to expand multiuser diversity gain versus static frequency planning [33], while coordinated load balancing and intercell scheduling were investigated in [40, 119]. Multi-cell power control algorithms were proposed in [86, 87]. The use of multiple antennas to suppress OCI has also been investigated as a coordinated single-cell transmission strategy, mainly in the form of receive combining. Optimal signal combining for space diversity reception with cochannel interference in cellular networks was proposed in [152, 153]. In [67, 76], spatial interference cancellation with multiple receive antennas has been exploited in ad hoc networks, which bear some similarity to multicell networks. Receive combining, however, can be applicable mainly in the uplink, as there are usually multiple antennas at the BS but only a small number of antennas at the mobile. Downlink beamforming in multicell scenarios was investigated in [37, 45], with the objective of minimizing the transmit power to support required receive SINR constraints at mobiles.

5.2 Contributions

In this chapter, we investigate spatial ICIC using ZF precoding to suppress downlink OCI and improve the system throughput. The main contributions are

summarized as follows.

Throughput analysis and adaptive ICIC: We provide closed-form expressions for the ergodic achievable sum rates when BSs take different transmission strategies, including selfish beamforming and doing ICIC for some of the neighboring cells. Adaptive ICIC is proposed to maximize the sum throughput by jointly selecting the transmission strategy at each BS based on user locations.

Strategy selection: It is shown that when the edge SNR is high, each BS tends to do ICIC for neighboring cells; when the edge SNR is low, each BS tends to do beamforming for its own user without ICIC; for medium edge SNR, the proposed adaptive strategy improves the sum and edge throughput and also reduces the required CSI compared to static ICIC. Numerical results show that in a 3-cell network the average throughput is increased by about half while the edge throughput is increased three-fold when the average edge SNR is 15 dB. In addition, with the sum throughput as the performance metric, the BS with a cell interior user is willing to help the edge user in the neighboring cell, i.e. it encourages fairness.

Impact of limited feedback: If the CSI at each BS is obtained through limited feedback, the induced quantization error will degrade the performance of ICIC. We provide accurate approximations for the achievable throughput with limited feedback. It is shown that to keep a constant rate loss versus perfect CSI, the number of feedback bits to the neighboring helper BS needs to grow linearly with both the number of transmit antennas and the edge SNR (in dB). With a constraint on the total number of feedback bits, the performance can be improved by adaptively allocating the available feedback bits.

5.3 System Model

We consider a multicell wireless network, where each BS has N_t antennas and each mobile user has a single antenna. Each mobile is associated with a *home BS*, which is the closest one. Universal frequency reuse is assumed. An active mobile, i.e. the one being scheduled for transmission, receives a data signal from its home

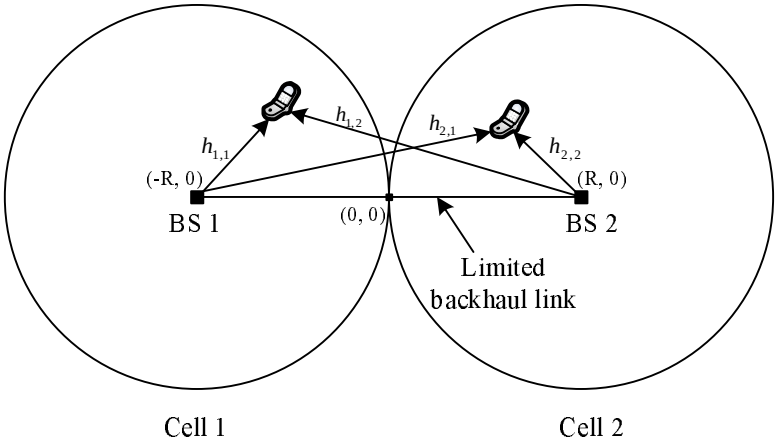


Figure 5.1: A two-cell network. Each BS is serving a home user, which is suffering OCI from the neighboring BS.

BS while suffering OCI from other BSs. ICIC in the spatial domain using multiple antennas is applied to suppress OCI. The BS applying ICIC for a user is called its *helper BS*. A 2-cell network is shown in Figure 5.1, which will be used as an instructive example in this section. We consider the downlink transmission, i.e. from the BS to mobiles. Following are some assumptions we make in our study.

Assumption 5. The neighboring BSs can exchange the location and CSI of each active user, but may not share traffic data.

With this assumption, each BS is able to do ICIC for its neighboring cells, but coordinated multicell transmission cannot be performed.

Assumption 6. There is one active user served in each cell at each time slot with precoding at the BS.

Denote the user and the BS in the i-th cell as the i-th user and the i-th BS, i.e. only a single user is active per BS per time slot, which precludes multi-user MIMO (MU-MIMO). The results could be extended to MU-MIMO in future work. With multi-antenna transmission at each BS, it is difficult to measure the interference from neighboring cells, which depends strongly on the active precoder, so we do not consider channel-dependent scheduling in the current work.

5.3.1 Adaptive Coordination

With multiple antennas, although each BS is able to do ICIC to cancel OCI for neighboring cells, this may be suboptimal, as ICIC will reduce the received signal power for its own user. For a 2-cell network, we assume each BS can select one of two strategies:

1. **Selfish beamforming**: it serves its own user with eigen-beamforming and does not cancel interference for the other cell. This strategy is denoted as BF.

2. **Interference cancellation**: it does interference cancellation for some of the neighboring cells. Denote $IC(\mathcal{I}_i)$ as the strategy that the i-th BS is doing ICIC for the users with indices in the set $\mathcal{I}_i$. In a 2-cell network, $IC(\mathcal{I}_i)$ is simplified as IC without ambiguity.

So the strategy set is $\mathcal{S}_1 = \{BF, IC\}$ and $\mathcal{S}_2 = \{BF, IC\}$ for BS 1 and BS 2, respectively, and the strategy pair taken by 2 BSs is $(s_1, s_2) \in \mathcal{S}_1 \times \mathcal{S}_2$, where $\mathcal{S}_1 \times \mathcal{S}_2$ is the Cartesian product.

When the active strategy pair is (s_1, s_2), the two received signals are given as

$$y_1(s_1, s_2) = \sqrt{P_{1,1}^r}\mathbf{h}_{1,1}^* \mathbf{f}_{1,s_1} x_1 + \sqrt{P_{1,2}^r}\mathbf{h}_{1,2}^* \mathbf{f}_{2,s_2} x_2 + z_1, \tag{5.1}$$

$$y_2(s_1, s_2) = \sqrt{P_{2,2}^r}\mathbf{h}_{2,2}^* \mathbf{f}_{2,s_2} x_2 + \sqrt{P_{2,1}^r}\mathbf{h}_{2,1}^* \mathbf{f}_{1,s_1} x_1 + z_2, \tag{5.2}$$

where $\mathbf{a}^*$ is the conjugate transpose of a vector $\mathbf{a}$ and

- $P_{i,j}^r$ is the received power at the i-th user from the j-th BS. We use the path loss model $P_{i,j}^r = P_0 \left(D_0/d_{i,j}\right)^\alpha$, where P_0 is the received signal power at the reference distance D_0, and $d_{i,j}$ is the distance between the user in the i-th cell and the j-th BS. In the following, we set $D_0 = R$, so P_0 is the average SNR at the cell edge. We assume equal transmit power at each BS, i.e. no power control is considered[1].

[1] Although power control can also be used to mitigate OCI and improve the system throughput [51], the emphasis in this chapter is on ICIC.

- z_i is the complex white Gaussian noise with zero mean and unit variance, i.e. $z_i \sim \mathcal{CN}(0,1)$. For a general multicell network, it may include interference from distant BSs.

- $\mathbf{h}_{i,j}$ is the $N_t \times 1$ channel vector from the j-th BS to the i-th user. We assume uncorrelated Rayleigh fading, so each component of $\mathbf{h}_{i,j}$ is i.i.d. $\mathcal{CN}(0,1)$.

- $\mathbf{f}_{i,s_i}$ is the precoding vector for the i-th user when BS i takes strategy s_i, $i = 1, 2$. It is normalized, i.e. $\|\mathbf{f}_{i,s_i}\|^2 = 1$, and its design will be discussed later in this section.

- x_i is the transmit signal for the i-th user, with the power constraint $\mathbb{E}[|x_i|^2] = 1$ for $i = 1, 2$.

The first term on the right hand side of (5.1) and (5.2) is the information signal, while the second term is the OCI. Taking user 1 as an example, the received signal-to-interference-plus-noise ratio (SINR) is

$$\mathrm{SINR}_1(s_1, s_2) = \frac{P^r_{1,1}|\mathbf{h}^*_{1,1}\mathbf{f}_{1,s_1}|^2}{1 + P^r_{1,2}|\mathbf{h}^*_{1,2}\mathbf{f}_{2,s_2}|^2}. \tag{5.3}$$

The achievable ergodic rate is

$$R_1(s_1, s_2) = \mathbb{E}\left[\log_2\left(1 + \mathrm{SINR}_1(s_1, s_2)\right)\right], \tag{5.4}$$

where $\mathbb{E}[\cdot]$ is the expectation operator.

The objective of our design is to select the strategy s_i for each BS to maximize the sum throughput, i.e. to solve the following problem

$$(s_1^*, s_2^*) = \arg \max_{s_1 \in \mathcal{S}_1, s_2 \in \mathcal{S}_2} R_1(s_1, s_2) + R_2(s_1, s_2). \tag{5.5}$$

This is called *adaptive ICIC*, and from (5.3) and (5.4) the adaptation is based on the locations of active users, which determine $P^r_{i,j}$, $i, j = 1, 2$. Therefore, BSs need to exchange user locations, but instantaneous CSI of a neighboring user is needed only when ICIC is applied to suppress OCI for this user. To solve the problem in (5.5) we need to first calculate the achievable sum throughput for different (s_1, s_2), which will be provided in Section 5.4.

Remark 12. *Although we use the sum throughput as the performance metric, our analysis can be easily extended to maximize a weighted sum throughput. In addition, in the following analysis and simulation, we will show that somewhat atypically, maximizing the sum throughput inherently provides fairness, and the proposed adaptive coordination strategy increases both the sum throughput and the edge throughput.*

5.3.2 Transmission Strategies

In this subsection, we describe the precoder design for different transmission strategies.

Eigen-beamforming In the single-cell scenario, eigen-beamforming is optimal for the MISO system with multiple transmit and a single receive antenna [136], for which the precoding vector is the channel direction, i.e. for the i-th user $\mathbf{f}_{i,BF} = \mathbf{h}_{i,i}/\|\mathbf{h}_{i,i}\|$. Therefore, the signal term is distributed as $|\mathbf{f}_{i,BF}^{*}\mathbf{h}_{i,i}|^2 \sim \chi_{2N_t}^2$, where χ_n^2 denotes the chi-square random variable (RV) with n degrees of freedom.

ICIC through ZF precoding With N_t antennas each BS can maximally precancel interference for up to $N_t - 1$ neighboring cells with ZF precoding. Taking cell 1 as an example, to cancel its interference for users in cell 2, 3, $\cdots$, K, $(K \leq N_t)$, the precoding vector $\mathbf{f}_{1,IC}$ needs to satisfy the orthogonality condition $\mathbf{f}_{1,IC}^{*}\mathbf{h}_{i,1} = 0$, for $i = 2, 3, \cdots, K$. Meanwhile, we also want to maximize the desired signal power $|\mathbf{f}_{1,IC}^{*}\mathbf{h}_{1,1}|^2$. This corresponds to choosing the precoding vector $\mathbf{f}_{1,IC}$ in the direction of the projection of vector $\mathbf{h}_{1,1}$ on the nullspace of vectors $\hat{\mathbf{H}} = [\mathbf{h}_{2,1}, \mathbf{h}_{3,1}, \cdots, \mathbf{h}_{K,1}]$ [76], i.e. the precoding vector is the normalized version of the following vector

$$\mathbf{w}_1^{(1)} = \left(\mathbf{I} - P_{\hat{\mathbf{H}}}\right)\mathbf{h}_1^{(1)}, \tag{5.6}$$

where $P_{\hat{\mathbf{H}}}$ is the projection on $\hat{\mathbf{H}}$, given as $P_{\hat{\mathbf{H}}} = \hat{\mathbf{H}}\left(\hat{\mathbf{H}}^{*}\hat{\mathbf{H}}\right)^{-1}\hat{\mathbf{H}}^{*}$. From [76], we have the distribution of the signal power as $|\mathbf{f}_{1,IC}^{*}\mathbf{h}_{1,1}|^2 \sim \chi_{2(N_t-(K-1))}^2$. This ICIC strategy with ZF precoding is low complex and provides closed-form analytical results. Although MMSE precoding outperforms ZF precoding at low SNR [113], as we will

show later that no ICIC is required when edge SNR is low. So there is negligible performance loss associated with applying ZF precoding instead of MMSE precoding.

Signal Power and Interference Power As shown in (5.3) and (5.4), the achievable throughput depends on the distributions of signal and interference terms. From the precoder design, we see that the received signal term of each user is a chi-square RV, with degrees of freedom depending on the transmission strategy of its home BS. For the interference power at the i-th user from the j-th BS, for $i \neq j$, if $s_j = IC(\mathcal{I}_j)$ and $i \in \mathcal{I}_j$, i.e. BS j does ICIC for the i-th user, then user i does not suffer interference from BS j; otherwise, the i-th user suffers interference distributed as $|\mathbf{f}_{j,s_j}^* \mathbf{h}_{i,j}|^2 \sim \chi_2^2$, which is because the design of the precoder $\mathbf{f}_{j,s_j}$ is independent of $\mathbf{h}_{i,j}$ and $|\mathbf{f}_{j,s_j}|^2 = 1$. Therefore, we have the following lemma on the distribution of the received signal and interference power.

Lemma 8. *The received signal power of the i-th user is distributed as*

$$|\mathbf{f}_{i,s_i}^* \mathbf{h}_{i,i}|^2 \sim \begin{cases} \chi_{2N_t}^2 & s_i = BF \\ \chi_{2(N_t-m)}^2 & s_i = IC(\mathcal{I}_i), |\mathcal{I}_i| = m, \end{cases} \tag{5.7}$$

where $|\mathcal{I}|$ is the cardinality of the set $\mathcal{I}$.

The interference power of the i-th user from the j-th BS is distributed as

$$|\mathbf{f}_{j,s_j}^* \mathbf{h}_{i,j}|^2 \begin{cases} = 0 & s_j = IC(\mathcal{I}_j), i \in \mathcal{I}_j \\ \sim \chi_2^2 & \text{otherwise.} \end{cases} \tag{5.8}$$

Remark 13. *From this lemma, we see that if one BS does interference cancellation for m neighboring cells instead of doing selfish beamforming, the received signal power of its own user changes from a $\chi_{2N_t}^2$ RV to a $\chi_{2(N_t-m)}^2$ RV, with the number of degrees of freedom reduced by $2m$; meanwhile, for the user in the neighboring cell helped by this BS, the interference power is reduced from a χ_2^2 RV to 0. The net effect on the sum throughput, however, is not clear. This is the focus in the following sections, i.e. to characterize the achievable sum throughput when BSs take different transmission strategies.*

5.4 Performance Analysis of a 2-cell Network

In this section, we focus on the 2-cell network depicted in Figure 5.1. We first derive the ergodic achievable throughput with different transmission strategy pairs (s_1, s_2) at two BSs, which are closed-form expressions and can be used to select (s_1, s_2) to maximize the sum throughput. Then we provide some insights on the transmission strategy selection.

5.4.1 Auxiliary Results

In this subsection, we provide two lemmas that will be used in the throughput analysis.

Lemma 9. *Assuming the RV X with distribution $X \sim \chi^2_{2M}$, we have*

$$R_{BF}(\gamma, M) = \mathbb{E}_X \left[\log_2 (1 + \gamma X)\right] = \log_2(e) e^{1/\gamma} \sum_{k=0}^{M-1} \frac{\Gamma(-k, 1/\gamma)}{\gamma^k}. \tag{5.9}$$

Proof. This result is provided as eq. (40) in [10]. □

Lemma 10. *Denote*

$$X \triangleq \frac{\gamma_1 Z}{1 + \gamma_2 Y},$$

where the RVs $Z \sim \chi^2_{2M}$, $Y \sim \chi^2_2$, and Z is independent of Y. Then

$$\begin{aligned} R_I^{(2)}(\gamma_1, \gamma_2, M) &= \mathbb{E}_X \left[\log_2(1 + X)\right] \\ &= \log_2(e) \sum_{i=0}^{M-1} \sum_{l=0}^{i} \frac{\gamma_1^{l+1-i}}{\gamma_2 (i-l)!} \cdot \hat{I}_3 \left(\frac{1}{\gamma_1}, \frac{\gamma_1}{\gamma_2}, i, l+1 \right), \end{aligned} \tag{5.10}$$

where $\hat{I}_3$ is the integral given in (3.22).

Proof. See Appendix 5.9.1. □

5.4.2 Throughput Analysis

Without loss of generality, we analyze the ergodic achievable throughput of user 1. In this part, we consider perfect CSI at the BS. The main result is given in the following theorem.

Theorem 8. *The ergodic achievable throughput of user 1 in a 2-cell network with given user locations and perfect CSI is given by*

$$R_1(s_1,s_2)=\begin{cases} R_I^{(2)}(P_{1,1}^r,P_{1,2}^r,N_t) & (s_1,s_2)=(BF,BF) \\ R_{BF}(P_{1,1}^r,N_t) & (s_1,s_2)=(BF,IC) \\ R_{BF}(P_{1,1}^r,N_t-1) & (s_1,s_2)=(IC,IC) \\ R_I^{(2)}(P_{1,1}^r,P_{1,2}^r,N_t-1) & (s_1,s_2)=(IC,BF) \end{cases} \tag{5.11}$$

where R_{BF} and $R_I^{(2)}$ are given in (5.9) *and* (5.10)*, respectively.*

Proof. The results are from *Lemma 9* for $s_2 = IC$ and *Lemma 10* for $s_2 = BF$, together with *Lemma 8*. □

The results in *Theorem 8* are closed-form expressions, from which we are able to select the strategy pair to maximize the sum throughput. However, the expressions in (5.9) and (5.10) are complicated and provide little insight. In the following, we provide a heuristic discussion on the strategy selection for different interference-to-noise ratio (INR) scenarios.

Both users are noise-limited: This scenario corresponds to $\text{INR}_1 \ll 1$ and $\text{INR}_2 \ll 1$. It may happen when both users are in the cell interior, or when the edge SNR is very low. For this scenario, as noise dominates OCI, ICIC provides a marginal gain, and each BS is willing to do beamforming to increase the received signal power for its own user, i.e. the strategy pair will be (BF, BF). Therefore, *there is no need to do ICIC in this scenario.*

Both users are interference-limited: This scenario corresponds to $\text{INR}_1 \gg 1$ and $\text{INR}_2 \gg 1$. This may happen when both users are at the cell edge and the transmit power is relatively high compared to the additive noise. As users suffer a higher level of OCI in this scenario, the BS will do ICIC for the neighboring cell to increase the sum throughput, i.e. the strategy pair will be (IC, IC).

One user is noise-limited, and the other is interference-limited: This scenario corresponds to $\text{INR}_1 \ll 1$ and $\text{INR}_2 \gg 1$. This may happen when user 1 is in the cell interior and user 2 is at the cell edge. For the interior user, it normally enjoys

a high SINR, so its throughput is limited by bandwidth. This means that doing ICIC for user 2 will not hurt user 1 so much, as the received signal power reduction for user 1 only brings a throughput loss in a log scale. On the other hand, user 2 is limited by OCI, so it requires ICIC from BS 1. Meanwhile, BS 2 will do beamforming for user 2 to increase the signal power, as the throughput of user 2 is power-limited. Therefore, the strategy pair will be (IC, BF).

Remark 14. *Although this is just a heuristic discussion, it shows that different strategy pairs will be selected for different scenarios, depending on user locations and average edge SNR. The ICIC strategy is not always necessary. The third scenario is of particular interest, as it shows that even with sum throughput as the metric the BS with an interior user (high rate) is willing to help the edge user (low rate) in the neighboring cell, i.e.* encouraging fairness. *Note that the strategy pair selection in the above discussion may not be the actual selection, and the actual strategy depends on user locations, the additive noise level and edge SNR, which can be determined from* (5.11).

In Figure 5.2, we compare the simulation and calculation results. Referring to Figure 5.1, user 1 is fixed at the cell edge $(-0.1R, 0)$, while user 2 is moving on the line connecting BS 1 and BS 2, with location $(x_2R, 0)$. We see that for average edge SNR $P_0 = 10$ dB, and for the considered locations, the strategy pair (IC, IC) is always selected. In Figure 5.3, we plot the selected strategy pairs for different user locations, where user 1 and 2 are moving on the line connecting BS 1 and BS 2. The x- and y-axis are the distance for user 1 and user 2 from the central point $(0, 0)$, respectively. The following observations can be made:

1. When the edge SNR is small ($P_0 = -5$ dB), (BF, BF) dominates, as the throughput is limited by noise and each BS tries to increase the received signal power for its own user.

2. When the edge SNR is large ($P_0 = 10$ dB), (IC, IC) dominates, as the throughput is limited by OCI and each BS does ICIC for neighboring cells.

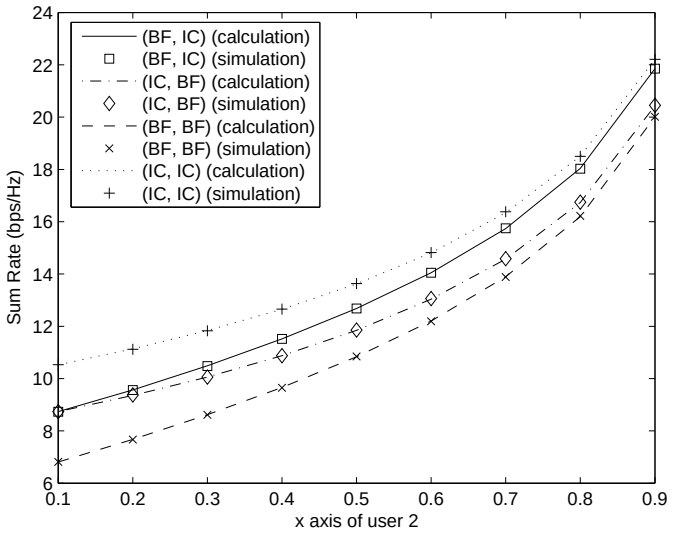


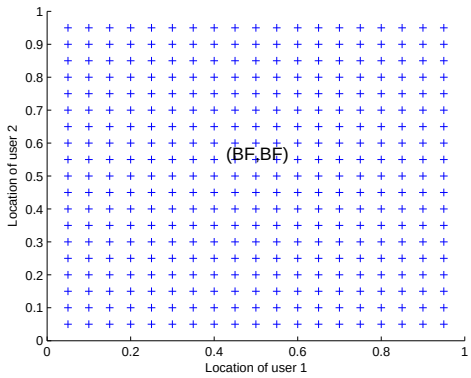
Figure 5.2: Simulation and calculation results for different transmission strategy pairs. User 1 is at the cell edge $(-0.1R, 0)$, and user 2 is moving from the cell edge to cell interior, $P_0 = 10$ dB, $\alpha = 3.7$, $N_t = 4$.

3. For medium SNR ($P_0 = 5$ dB), the selected strategy pair depends on the user locations. Specifically, it shows that when both users are in cell interior, i.e. INRs are small, (BF, BF) is selected; when both users are at cell edge, i.e. INRs are large, (IC, IC) is selected; when one user is in cell interior, and the other is at cell edge, the BS with the interior user will do ICIC for the edge user.

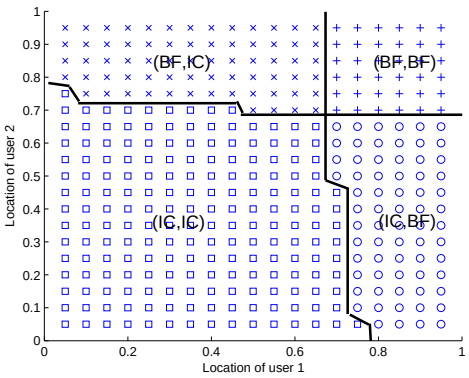
These observations agree with the above discussion and motivate to adaptively select transmission strategies.

5.5 From 3-cell to Multicell Networks

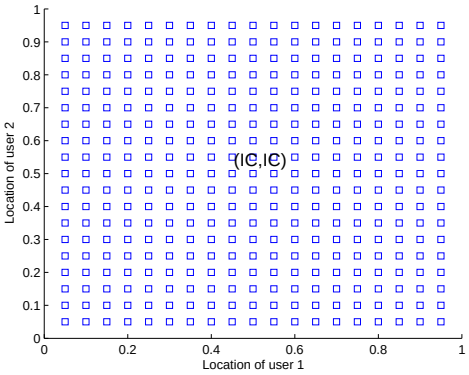
The investigation of the simplified 2-cell network provided insights about the strategy selection and motivated the adaptive coordination, but the result cannot be readily implemented in a general multicell network. In this section, we first extend our adaptive strategy to a 3-cell network. We derive closed-form expressions for the achievable throughput for the strategy selection. Based on the results for 3-cell networks, we also propose approaches to extend the adaptive coordination to general



(a) $P_0 = -5$ dB



(b) $P_0 = 5$ dB



(c) $P_0 = 10$ dB

Figure 5.3: Selected transmission strategies for different user locations in a 2-cell network, where $\alpha = 3.7$, $N_t = 4$, user 1 and user 2 are on the line connecting BS 1 and BS 2. The mark 'x' denotes $(s_1, s_2) = (BF, IC)$, 'o' denotes $(s_1, s_2) = (IC, BF)$, '+' denotes $(s_1, s_2) = (BF, BF)$, and '□' denotes $(s_1, s_2) = (IC, IC)$.

multicell networks.

5.5.1 The Strategy Set

With 3 cells coordinating with each other, each BS has four different strategies. Taking user 1 as an example, we describe different strategies as follows.

1. **Selfish beamforming:** BS 1 does beamforming for user 1, denoted as $s_1 = BF$.
2. **ICIC for 2 neighboring cells:** BS 1 does ICIC for both cell 2 and 3, which requires $N_t \geq 3$. This is denoted as $s_1 = IC(\{2,3\})$.
3. **ICIC for cell 2:** BS 1 does ICIC for cell 2, denoted as $s_1 = IC(2)$.
4. **ICIC for cell 3:** BS 1 does ICIC for cell 3, denoted as $s_1 = IC(3)$.

To reduce the size of the strategy set, we combine strategy 3 and 4 as a single strategy, for which BS 1 does ICIC for the neighboring cell that suffers a higher level of average OCI from BS 1, i.e. to help the neighboring cell user that is closer to BS 1. This is a reasonable approach and reduces the complexity of the strategy selection process.

Therefore, the strategy set for user 1 is $\bar{\mathbb{S}}_1 = \{BF, IC(2 \text{ or } 3), IC(\{2,3\})\}$. There are a total of $3^3 = 27$ different strategy combinations for 3 users, $(s_1, s_2, s_3) \in \bar{\mathbb{S}}_1 \times \bar{\mathbb{S}}_2 \times \bar{\mathbb{S}}_3$.

5.5.2 Throughput Analysis

First, we present the following lemma for throughput analysis.

Lemma 11. *Denote*

$$X \triangleq \frac{\alpha Z}{1 + \delta_1 Y_1 + \delta_2 Y_2}, \tag{5.12}$$

where $Z \sim \chi^2_{2M}$, $Y_1 \sim \chi^2_2$, $Y_2 \sim \chi^2_2$, *and they are independent. Then*

$$\begin{aligned} &R_I^{(3)}(\alpha, \delta_1, \delta_2, M) = \mathbb{E}_X\left[\log_2(1+X)\right] \\ =&\log_2(e) \sum_{i=0}^{M-1} \sum_{l=0}^{i} \frac{\alpha^{l-i+1}}{(\delta_1 - \delta_2)(i-l)!} \left[\hat{I}_3\left(\frac{1}{\alpha}, \frac{\alpha}{\delta_1}, i, l+1\right) - \hat{I}_3\left(\frac{1}{\alpha}, \frac{\alpha}{\delta_2}, i, l+1\right)\right], \end{aligned} \tag{5.13}$$

where $\hat{I}_3(\cdot,\cdot,\cdot,\cdot)$ *is the integral given in* (3.22).

Proof. The proof is similar to the one in Appendix 5.9.1 for *Lemma 10.* □

Taking the first user as an example, its received SINR with the strategy $\mathbf{s} = (s_1, s_2, s_3)$ is

$$\mathrm{SINR}_1(\mathbf{s}) = \frac{P^r_{1,1}|\mathbf{h}^*_{1,1}\mathbf{f}_{1,s_1}|^2}{1 + P^r_{1,2}|\mathbf{h}^*_{1,2}\mathbf{f}_{2,s_2}|^2 + P^r_{1,3}|\mathbf{h}^*_{1,3}\mathbf{f}_{3,s_3}|^2}. \tag{5.14}$$

The achievable rate is

$$R_1(\mathbf{s}) = \mathbb{E}\left[\log_2\left(1 + \mathrm{SINR}_1(s_1, s_2, s_3)\right)\right], \tag{5.15}$$

for which a closed-form expression is given in the following theorem.

Theorem 9. *The ergodic achievable throughput of user 1 in a 3-cell network with given user locations and perfect CSI is given by*

$$R_1(\mathbf{s}) = \begin{cases} R_I^{(3)}(P^r_{1,1}, P^r_{1,2}, P^r_{1,3}, M) & s_2 = BF, s_3 = BF \\ R_I^{(2)}(P^r_{1,1}, P^r_{1,j}, M) & s_j = IC(\mathcal{I}_j), 1 \in \mathcal{I}_j, j = 2 \text{ or } 3 \\ R_{BF}(P^r_{1,1}, M) & s_j = IC(\mathcal{I}_j), 1 \in \mathcal{I}_j, j = 2, 3 \end{cases} \tag{5.16}$$

where R_{BF}, $R_I^{(2)}$, *and* $R_I^{(3)}$ *are given in* (5.9), (5.10), *and* (5.13) *respectively. The parameter* M *depends on the distribution of the signal term, which subsequently depends on* s_1*:*

$$M = \begin{cases} N_t & s_1 = BF \\ N_t - m & s_1 = IC(\mathcal{I}_1), |\mathcal{I}_1| = m. \end{cases} \tag{5.17}$$

Proof. The results come from *Lemma 8* and *Lemma 11.* □

Based on this theorem, we are able to select the transmission strategy at each BS to maximize the sum throughput. Note that the strategy selection is in a coordinated way, i.e. the 3 BSs jointly determine the set (s_1, s_2, s_3), as the objective function is common for all the BSs. It explicitly assumes that each BS knows the strategy taken by other BSs.

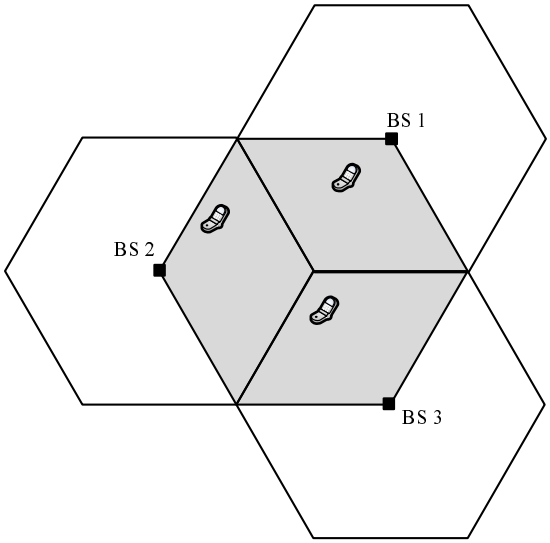


Figure 5.4: A 3-cell network. The shadow area is considered as the "inner area", where users suffer high OCI from neighboring cells.

5.5.3 Extension to Multicell Networks

In this subsection, we propose approaches to extend our results to a general multicell setting. A detailed investigation is beyond the scope of this dissertation but is feasible in principle.

One approach is to apply the proposed adaptive ICIC strategy with cell sectoring, as in [159]. By using 120-degree sectoring in each cell, every 3 neighboring cells can coordinate with each other to serve users in the shadow area shown in Figure 5.4, where the 3 BSs jointly select the transmission strategy based on the results in *Theorem 9*.

It is also possible to implement the adaptive ICIC strategy in a distributed way. In this approach, each BS determines its transmission strategy independently rather than in a coordinated way. The main idea is for each BS to select its transmission strategy by itself. To do this, each BS needs to estimate if there is a sum throughput gain by providing ICIC for its neighboring cells. If the sum throughput is increased, it will select ICIC as its strategy; otherwise, it will perform beamforming for its own user. As each user is located in the interior area of a certain 3-cell sub-network, as in

the shadow area in Figure 5.4, its achievable throughput can be estimated based on (5.16) by approximating the interference from outer cells as white Gaussian noise. No cluster structure is used, so this approach can be adopted in a network of an arbitrary size.

5.6 Impact of Limited Feedback

We have assumed perfect CSI at the BS in the results provided thus far. However, in realistic scenarios, there will always be inaccuracy in the available CSI. In this section, we consider a FDD (Frequency Division Duplex) system where CSI is obtained through limited feedback [97]. As limited feedback sends quantized channel information to the transmitter, it introduces quantization error to the available CSI. We will analyze the impact of limited feedback, and consider feedback design for adaptive ICIC transmission.

5.6.1 Limited Feedback

With limited feedback, the channel direction information (CDI) is fed back using a quantization codebook known at both the transmitter and receiver. The quantization is chosen from a codebook of unit norm vectors of size $L = 2^B$, where B is the number of feedback bits. Denote the codebook as $\mathcal{C} = \{\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_L\}$. Each user quantizes its channel direction to the closest codeword, measured by the inner product. Therefore, the quantized channel direction is

$$\hat{\mathbf{h}}_{i,j} = \arg\max_{\mathbf{c}\in\mathcal{C}} |\tilde{\mathbf{h}}_{i,j}^{*}\mathbf{c}|, \tag{5.18}$$

where $\tilde{\mathbf{h}}_{i,j} = \frac{\mathbf{h}_{i,j}}{\|\mathbf{h}_{i,j}\|}$ is the actual channel direction. Then each user feeds back B bits to indicate the index of this codeword in the codebook $\mathcal{C}$. We assume the channel estimation at each user is perfect, and the feedback channel is error-free and without delay. Random vector quantization (RVQ) [75, 120] is used to facilitate the analysis, where each quantization vector is independently chosen from the isotropic distribution on the N_t-dimensional unit sphere.

If ICIC is performed for the i-th user by some of its neighboring BSs, this user needs to estimate channel directions from multiple BSs, which are then independently quantized and fed back to its home BS. Then the home BS can forward the associated CDI to neighboring BSs through backhaul connection.

Assumption 7. *The i-th user uses the codebook $\mathcal{C}_{i,j}$ to quantize CDI for the j-th BS, which is of size $L_{i,j} = 2^{B_{i,j}}$. If $L_{i,j}$ is the same for different j, user i can use the same quantization codebook, but the codebooks are different from user to user.*

As will be shown later, the quantization for channel directions of different BSs have different impacts on the system performance, so different $L_{i,j}$ for different i and j may provide better performance. Different users employing different codebooks is to avoid the same quantized CDI from multiple users at the same BS.

5.6.2 Throughput Analysis

First, we consider the statistics of the quantized CDI. Let $\cos\theta_{i,j} = |\tilde{\mathbf{h}}_{i,j}^* \hat{\mathbf{h}}_{i,j}|$, where $\theta_{i,j} = \angle\left(\tilde{\mathbf{h}}_{i,j}, \hat{\mathbf{h}}_{i,j}\right)$, then we have [13]

$$\xi_{i,j} = \mathbb{E}_{\theta_{i,j}}\left[\cos^2\theta_{i,j}\right] = 1 - L_{i,j} \cdot \beta\left(L_{i,j}, \frac{N_t}{N_t - 1}\right), \tag{5.19}$$

where $\beta(x, y)$ is the Beta function, i.e. $\beta(x, y) = \frac{\Gamma(x)\Gamma(y)}{\gamma(x+y)}$ with $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\mathrm{d}t$ as the Gamma function.

To investigate the impact of limited feedback, we first analyze the received signal power and interference power with limited feedback.

Lemma 12. *If CDI at the BS is obtained through limited feedback, the received signal power of the i-th user with the expectation on $\theta_{i,i}$ can be approximated as $\mathbb{E}_{\theta_{i,i}}\left[|\mathbf{h}_{i,i}^* \mathbf{f}_{i,s_i}|^2\right] \approx \xi_{i,i} X$, where $\xi_{i,i}$ is given in* (5.19) *and the RV X is distributed as*

$$X \sim \begin{cases} \chi^2_{2N_t} & s_i = BF \\ \chi^2_{2(N_t-m)} & s_i = IC(\mathcal{I}_i), |\mathcal{I}_i| = m. \end{cases} \tag{5.20}$$

The interference power of the i-th user from the j-th BS is distributed as

$$|\mathbf{h}_{i,j}^* \mathbf{f}_{j,s_j}|^2 \sim \begin{cases} \kappa_{i,j}\chi_2^2 & s_j = IC(\mathcal{I}_j), i \in \mathcal{I}_j \\ \chi_2^2 & \textit{otherwise}, \end{cases} \tag{5.21}$$

where $\kappa_{i,j} = 2^{-\frac{B_{i,j}}{N_t - 1}}$.

Proof. See Appendix 5.9.2. □

Remark 15. *From this lemma, we see that limited feedback has differing impact on the received signal term and the interference term: it only changes the mean, not the distribution of the signal term; for the interference term, the distribution is the same without ICIC, but limited feedback causes residual interference with ICIC. In addition, at high edge SNR, the impact of limited feedback on the signal term only causes a constant rate loss of* $\log \xi_{i,i}$ *for the i-th user, but the resulting residual OCI increases with edge SNR and limits the system throughput. Therefore, the CDI need not be of the same accuracy for the home BS and the helper BS, which leaves flexibility for the feedback design.*

Based on the above lemma, we provide the following theorem on the achievable throughput with limited feedback.

Theorem 10. *The achievable throughput of user 1 in a 3-cell network with given user locations and limited feedback is approximated by*

$$R_1(\mathbf{s}) \approx \begin{cases} R_I^{(3)}(\xi_{1,1}P_{1,1}^r, P_{1,2}^r, P_{1,3}^r, M) & s_2 = BF, s_3 = BF \\ R_I^{(3)}(\xi_{1,1}P_{1,1}^r, P_{1,j}^r, \kappa_{1,k}P_{1,k}^r, M) & s_j = IC(\mathfrak{I}_j), 1 \in \mathfrak{I}_j, j = 2 \text{ or } 3 \\ R_I^{(3)}(\xi_{1,1}P_{1,1}^r, \kappa_{1,2}P_{1,2}^r, \kappa_{1,3}P_{1,3}^r M) & s_j = IC(\mathfrak{I}_j), 1 \in \mathfrak{I}_j, j = 2, 3 \end{cases} \quad (5.22)$$

where $R_I^{(3)}$ *is given in* (5.13)*, and* M *is given by* (5.17)*.*

Proof. See Appendix 5.9.3. □

Remark 16. *This result can be easily modified for a 2-cell network. Note that with limited feedback, each user always suffers from OCI, due to co-channel transmission and/or imperfect interference cancellation. Therefore, the transmission strategy selection now depends not only on user locations, average edge SNR, but also on the number of feedback bits.*

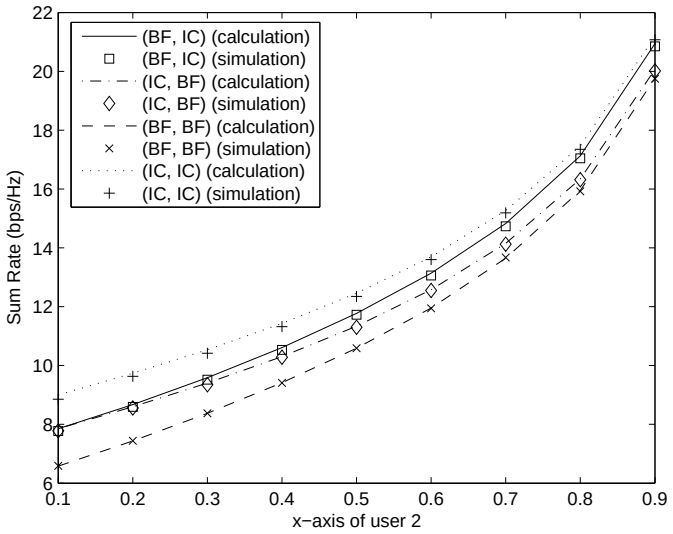


Figure 5.5: Simulations and approximations for the sum throughput in a 2-cell network. User 2 is at $(-.1R, 0)$, $P_0 = 10$ dB, $\alpha = 3.7$, $B = 10$, and $N_t = 4$.

In Figure 5.5, we compare the simulation and approximation results with limited feedback in the same setting as Figure 5.2 and with feedback bits for each channel direction to be $B = 10$. We see that the approximations are very accurate. Compared to Figure 5.2, the performance gain due to ICIC is reduced, but the strategy pair (IC, IC) is still preferred. We can also get a similar plot as Figure 5.3, which is omitted due to space limitation, but similar observations can be made except that operating regions with the ICIC strategy shrink.

5.6.3 Limited Feedback Design

With ICIC, each user needs to feed back multiple channel directions. The feedback should be carefully designed as the resource on the feedback channel is limited. In this subsection, we consider feedback in the following two scenarios:

- If the number of feedback bits can be varying, how many bits do we need to keep a constant rate loss versus the perfect CSI case?
- If the total number of feedback bits is fixed, how should we allocate them between

the CSI feedback for the home BS and the CSI feedback for the helper BS?

5.6.3.1 Feedback bits for a constant rate loss

If we can vary the number of feedback bits, based on the rate loss analysis, we provide the following theorem on the required scaling of feedback bits with different system parameters to keep a constant rate loss.

Theorem 11. *In a 3-cell network, to keep a constant rate loss of* $\log_2 \delta_R$ *bps/Hz compared to perfect CSI, the number of feedback bits for each helper BS needs to satisfy*

$$B^\star \geq (N_t - 1) \log_2 \left(\frac{2P_0}{\delta_R - 1} \right). \tag{5.23}$$

Proof. See Appendix 5.9.4. □

We see that similar to multiuser MIMO systems [75, 169], the feedback bit rate needs to increase linearly with both N_t and P_0 (in dB). The difference is that for ICIC there is no such requirement on the feedback of the CSI to the home BS, as it provides a fixed rate loss with a fixed number of feedback bits at high SNR according to *Lemma 12*, so we do not to increase the feedback bits for this link.

5.6.3.2 Feedback bits allocation

As shown in *Lemma 12*, the CSI accuracy for the home BS and the helper BS has different impact on the performance. This indicates that with a fixed number of feedback bits it is possible to improve the performance by adaptively allocating the total feedback bits for the home BS and the helper BS rather than an equal allocation.

With feedback bits allocation, the maximum achievable throughput for a given scenario is now given as

$$R_{\text{sum}} = \max_{\mathbf{s} \in \mathcal{S}, \sum_j B_{i,j} = B} \sum_{i=1}^{3} R_i(\mathbf{s}, B_{i,1}, B_{i,2}, B_{i,3}), \tag{5.24}$$

where $\mathbf{s} = (s_1, s_2, s_3)$ and the expression for R_i is given in (5.22). This is a combinatorial optimization problem, but may be solved by an exhaustive search for small

B. To reduce the search space, we can add additional constraints such as forcing the number of feedback bits to be the same for all the helper BSs.

5.7 Numerical Results

In this section, we present some numerical results to show the performance of our proposed adaptive ICIC strategy and build intuition. A 3-cell network as shown in Figure 5.4 is considered, where there is one active user randomly located in each cell in the shadow area. The radius of each cell is $R = 1$ km, the path loss exponent is 3.7, and $N_t = 4$.

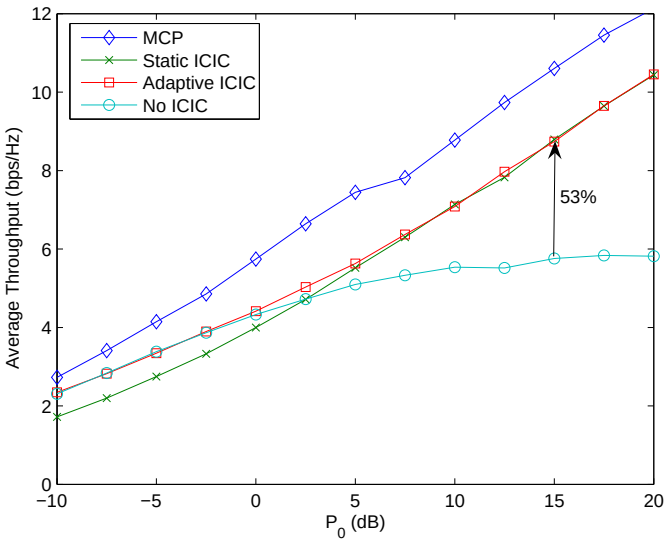
5.7.1 Performance comparison with perfect CSI

In this part, we compare the performance of four systems with different transmission strategies: the system with multiple-cell precoding with per-BS power constraint[2], i.e. all the three BSs jointly design precoders to serve the three users, denoted as *MCP*, the system without ICIC, i.e. each BS does selfish beamforming for its own user, denoted as *no ICIC*, the system where each BS always does ICIC for both of its neighboring cells, denoted as *static ICIC*, and the system with the proposed adaptive ICIC strategy where the transmission strategy at each BS is selected based on *Theorem 9*, denoted as *adaptive ICIC*.

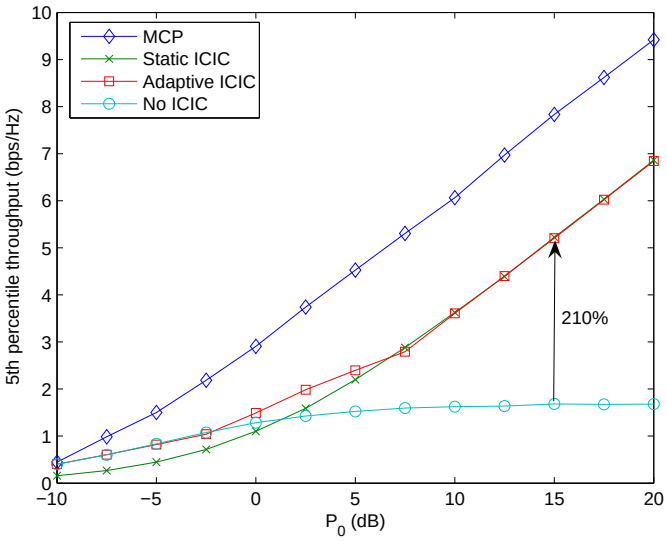
The MCP system serves as a benchmark, which is a coordinated multi-cell transmission strategy and requires full CSI and data sharing. Figure 5.8 shows the average throughput and the 5th percentile throughput, representing the cell edge throughput, for each of the four systems. We see that the performance difference depends on the edge SNR P_0:

For low P_0, the *static ICIC* system provides lower throughput than the other two, which shows there is no need to perform ICIC in this regime.

[2]The precoding is designed with zero-forcing criterion, and the power control with per-BS power constraint is similar to the one proposed in Section 4.4.2.1 for clustered linear precoding.



(a) Average throughput per cell



(b) 5th percentile throughput

Figure 5.6: Comparison of systems with different transmission strategies in a 3-cell network with perfect CSI.

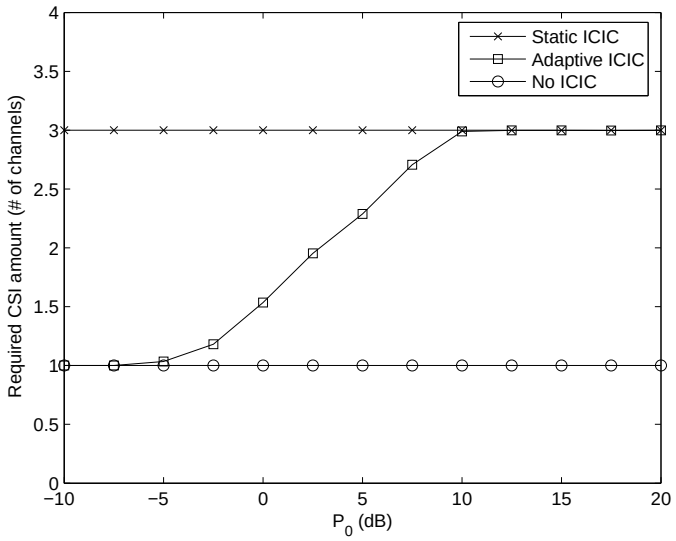


Figure 5.7: The CSI requirements for different systems, which are expressed in number of channel directions.

For high P_0, there is a rate ceiling for the *no ICIC* system, as its throughput is limited by OCI. The *static ICIC* and *adaptive ICIC* systems have similar performance, and both provide significant throughput gain over the *no ICIC* system, e.g for $P_0 = 15$ dB[3], the average throughput gain is 53% while the edge throughput gain is 210%.

For medium P_0 $(0 \sim 5)$, *adaptive ICIC* outperforms both *no ICIC* and *static ICIC*, which shows that we should not simply switch between selfish beamforming and static ICIC but should adaptively and jointly select the transmission technique at each BS.

Compared to *MCP*, *adaptive ICIC* achieves the same scaling with edge SNR of both the average and edge throughput, with a constant rate loss. The rate loss is about 2 bps/Hz and 2.5 bps/Hz for average throughput and edge throughput, respectively. For $P_0 = 15$ dB, *adaptive ICIC* achieves 62% and 71% of the throughput gain of *MCP* over the *no ICIC* system. Note that the proposed *adaptive ICIC*

[3]As shown in [65], an edge SNR$\sim$ 15 dB can be obtained with reasonable assumptions.

strategy does not require inter-BS data exchange, and is of lower complexity due to single-cell processing.

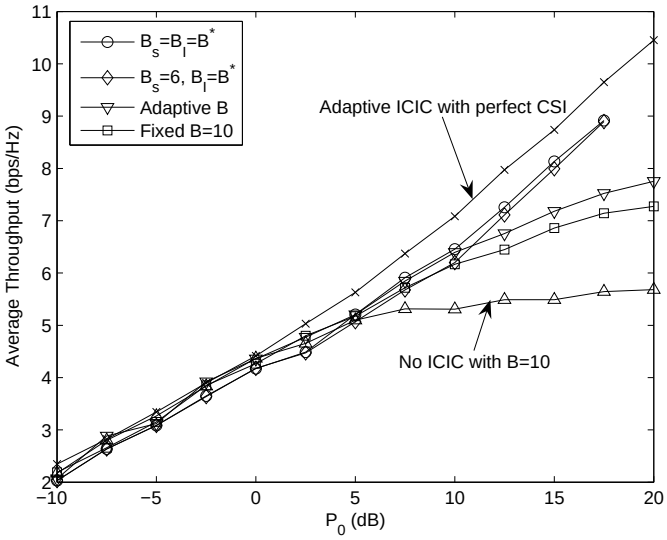
Compared to *static ICIC*, the *adaptive ICIC* system is able to reduce the amount of required CSI, as the CSI for a neighboring BS is needed only when this BS does ICIC for the user. Figure 5.7 compares the amount of CSI requirement for different systems, in number of channel directions. For medium P_0, *adaptive ICIC* reduces the required CSI amount compared to *static ICIC* while provides higher throughput than the other two systems.

5.7.2 Impact of limited feedback

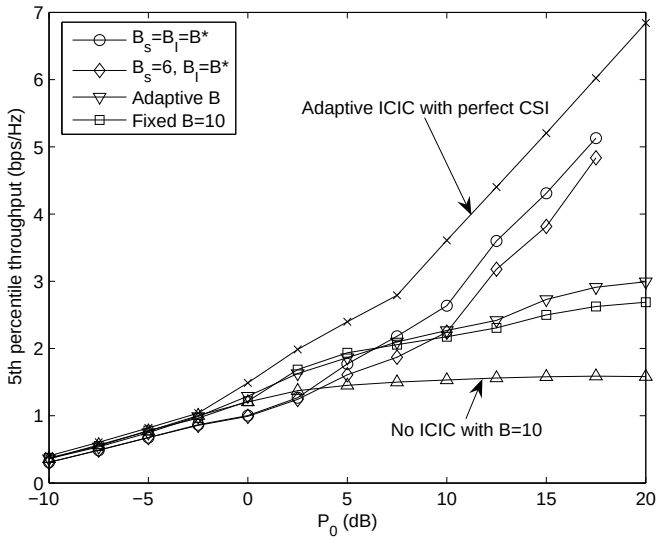
In Figure 5.8, we show the system performance when CSI is provided by limited feedback. We assume each user feeds back B_s bits for the CSI to its home BS and B_I bits for the CSI to each of its helper BSs if required. Adaptive ICIC based on (5.22) is applied.

First, we scale B_I according to (5.23) with the rate loss target $\delta_R = 1$ bps/Hz, i.e. $B_I = B^\star$, while B_s has the same scaling ($B_s = B^\star$) or is fixed to be $B_s = 6$. It shows that by scaling B_I, we can obtain a constant rate loss (less than 1 bps/Hz for both average and edge throughput) to the system with perfect CSI. In addition, we do not need to scale B_s, as a fixed B_s only causes a fixed rate loss at high edge SNR. Note that at high P_0, a large value of B_I is required, e.g. $B_I = 18$ for $P_0 = 15$.

Next, we assume the total number of feedback bits is fixed to be 30 bits, i.e. $B_s + 2B_I = 30$, and compare adaptive bit allocation according to (5.24) with uniform allocation ($B_s = B_I = 10$). For adaptive bit allocation, each user and BS need to maintain multiple codebooks, and to reduce the complexity we limit the allocation bit pair as $(B_s, B_I) \in \{(10, 10), (8, 11), (6, 12), (4, 13), (2, 14)\}$. We see that adaptive bit allocation provides better performance at high P_0, but the throughput gain over uniform allocation is marginal ($\sim 6\%$ for average throughput and $\sim 11\%$ for edge throughput). This is because the total number of feedback bits is not large enough. For example, to keep a rate loss of 1 bps/Hz to perfect CSI case, we need $B_I = 18$ for



(a) Average throughput per cell



(b) 5th percentile throughput

Figure 5.8: Comparison of systems with different feedback strategies in a 3-cell network, where B_s and B_I are the numbers of feedback bits for the home BS and the helper BS, respectively, and $B^\star$ is given in (5.23) for different P_0.

$P_0 = 15$, so $B_s + 2B_I > 36$, which can not be satisfied with the available number of bits. However, we see that all the adaptive ICIC systems provide a significant gain for both average and edge throughput over the system without ICIC even with limited feedback.

5.8 Conclusion

In this chapter, we investigated spatial ICIC to suppress OCI in a multicell wireless network. An adaptive strategy was proposed, where multiple BSs jointly select transmission strategies. ICIC is a type of coordinated single-cell transmission, so the system complexity is low and no central processing unit is required. In addition, it has a low overhead as only user locations are required for strategy selection, and instantaneous CSI is needed only at the home BS and the neighboring BS that does ICIC for the user. Numerical results showed that ICIC provides both average and edge throughput gain, and at medium edge SNR the adaptive strategy outperforms both the system without ICIC and the one with static ICIC. Even when the CSI is provided by limited feedback, the ICIC system still provides significant throughput gain with carefully designed feedback strategies. Given the consistent results for two and three cells, we conjecture that the results and design intuition extend to large cellular networks.

5.9 Appendix

5.9.1 Proof of Lemma 10

The proof is similar to the one for *Corollary 3* in Chapter 3, and in this appendix we only provide some key steps. First, the cumulative distribution function (cdf) of the RV X can be derived as

$$F_X(x) = 1 - \sum_{i=0}^{M-1} \sum_{l=0}^{i} \frac{\gamma_1^{l+1-i}}{\gamma_2 (i-l)!} \cdot \frac{x^i e^{-x/\gamma_1}}{\left(x + \frac{\gamma_1}{\gamma_2}\right)^{l+1}}. \tag{5.25}$$

The expectation of $\ln(1+x)$ on X is then derived as follows.

$$\begin{aligned}
&\mathbb{E}_X\left[\ln(1+X)\right] = \int_0^\infty \ln(1+x)\mathrm{d}F_X \\
&\overset{(a)}{=} \int_0^\infty \frac{1-F_X(x)}{x+1}\mathrm{d}x \\
&= \sum_{i=0}^{M-1}\sum_{l=0}^{i}\frac{\gamma_1^{l+1-i}}{\gamma_2(i-l)!}\int_0^\infty \frac{x^i e^{-x/\gamma_1}}{(x+1)\left(x+\frac{\gamma_1}{\gamma_2}\right)^{l+1}}\mathrm{d}x \\
&= \sum_{i=0}^{M-1}\sum_{l=0}^{i}\frac{\gamma_1^{l+1-i}}{\gamma_2(i-l)!}\cdot \hat{I}_3\left(\frac{1}{\gamma_1},\frac{\gamma_1}{\gamma_2},i,l+1\right),
\end{aligned} \tag{5.26}$$

where step (a) follows integration by parts and $\hat{I}_3(\cdot,\cdot,\cdot,\cdot)$ is the integral given in (3.22). Then we get (5.10).

5.9.2 Proof of Lemma 12

For the signal term, if $s_i = BF$, the beamforming vector is now based on the quantized channel direction, i.e. $\mathbf{f}_{i,BF} = \hat{\mathbf{h}}_{i,i}$. The signal term is

$$|\mathbf{h}_{i,i}^* \mathbf{f}_{i,BF}|^2 = |\mathbf{h}_{i,i}^* \hat{\mathbf{h}}_{i,i}|^2 = \cos^2\theta_{i,i}\|\mathbf{h}_{i,i}\|^2. \tag{5.27}$$

Taking the expectation on $\theta_{i,i}$, we have $\mathbb{E}_{\theta_{i,i}}\left[|\mathbf{h}_{i,1}^* \mathbf{f}_{i,BF}|^2\right] \sim \xi_{i,i}\chi^2_{2N_t}$.

If $s_i = IC(\mathcal{I}_i)$, $|\mathcal{I}_i| = m$, writing $\tilde{\mathbf{h}}_{i,i} = (\cos\theta_{i,i})\hat{\mathbf{h}}_{i,i} + (\sin\theta_{i,i})\mathbf{g}_{i,i}$, where $\mathbf{g}_{i,i}$ is orthogonal to $\hat{\mathbf{h}}_{i,i}$. The signal term with an expectation on $\theta_{i,i}$ is

$$\begin{aligned}
\mathbb{E}_{\theta_{i,i}}\left[|\mathbf{h}_{i,i}^* \mathbf{f}_{i,IC}|^2\right] &= \|\mathbf{h}_{i,i}\|^2 \cdot \mathbb{E}_{\theta_{i,i}}|(\cos\theta_{i,i})\hat{\mathbf{h}}_{i,i}^* \mathbf{f}_{i,IC} + (\sin\theta_{i,i})\mathbf{g}_{i,i}^* \mathbf{f}_{i,IC}|^2 \\
&\overset{(a)}{\approx} \|\mathbf{h}_{i,i}\|^2 \cdot \mathbb{E}_{\theta_{i,i}}|(\cos\theta_{i,i})\hat{\mathbf{h}}_{i,i}^* \mathbf{f}_{i,IC}|^2 \\
&= \mathbb{E}_{\theta_{i,i}}\left[\cos^2\theta_{i,i}\right] \cdot \|\mathbf{h}_{i,i}\|^2 |\hat{\mathbf{h}}_{i,i}^* \mathbf{f}_{i,IC}|^2 \\
&\overset{(b)}{\sim} \xi_{i,i}\chi^2_{2(N_t-m)}.
\end{aligned} \tag{5.28}$$

In step (a), we remove the $\sin\theta_{i,i}$ term, which is normally very small. The beamforming vector $\mathbf{f}_{i,IC}$ is in the direction of the projection of vector $\hat{\mathbf{h}}_{i,i}$ on the nullspace of $\hat{\mathbf{h}}_{j,i}$, $\forall j \neq i$, so similar to the perfect CSI case we have $\|\mathbf{h}_{i,i}\|^2|\hat{\mathbf{h}}_{i,i}^* \mathbf{f}_{i,IC}|^2 \sim \chi^2_{2(N_t-m)}$, which gives step (b).

For the interference power, take user 1 for an example, and consider the interference from BS 2. If BS 2 does not apply ICIC for user 1, then as in *Lemma 8*, $|\mathbf{h}_{1,2}^* \mathbf{f}_{2,IC}|^2 \sim \chi_2^2$. If BS 2 uses ICIC for user 1, with quantization error, there will be residual interference from BS 2. The interference power is $|\mathbf{h}_{1,2}^* \mathbf{f}_{2,IC}|^2$, where $\mathbf{f}_{2,IC}$ is in the direction of the projection of vector $\hat{\mathbf{h}}_{2,2}$ on the nullspace of $\hat{\mathbf{h}}_{1,2}$. Based on the quantization cell approximation, this interference term can be approximated as an exponential RV with mean $\kappa_{i,j} = 2^{-\frac{B_{i,j}}{N_t - 1}}$ [169], i.e. $|\mathbf{h}_{1,2}^* \mathbf{f}_{2,IC}|^2 \sim \kappa_{i,j} \chi_2^2$.

5.9.3 Proof of Theorem 10

In a 3-cell network, the achievable rate for user 1 is first approximated as

$$\begin{aligned} R_1(\mathbf{s}) &= \mathbb{E}_{\mathbf{h},\theta_{1,1}} \left[\log_2 \left(1 + \frac{P_{1,1}^r |\mathbf{h}_{1,1}^* \mathbf{f}_{1,s_1}|^2}{1 + P_{1,2}^r |\mathbf{h}_{1,2}^* \mathbf{f}_{2,s_2}|^2 + P_{1,3}^r |\mathbf{h}_{1,3}^* \mathbf{f}_{3,s_3}|^2} \right) \right] \\ &\approx \mathbb{E}_{\mathbf{h}} \left[\log_2 \left(1 + \frac{P_{1,1}^r \mathbb{E}_{\theta_{1,1}} \left[|\mathbf{h}_{1,1}^* \mathbf{f}_{1,s_1}|^2 \right]}{1 + P_{1,2}^r |\mathbf{h}_{1,2}^* \mathbf{f}_{2,s_2}|^2 + P_{1,3}^r |\mathbf{h}_{1,3}^* \mathbf{f}_{3,s_3}|^2} \right) \right]. \end{aligned} \tag{5.29}$$

Then based on *Lemma 11* and *Lemma 12*, we get the results in (5.22).

5.9.4 Proof of Theorem 11

At high edge SNR, each BS is doing ICIC for both its neighboring cells, and the system throughput is limited by the residual OCI. As the two neighboring cells are symmetric, let the number of feedback bits be B_I for each of them. As shown in [75, 169], the rate loss due to imperfect CSI is upper bounded as

$$\begin{aligned} \Delta R &\leq \mathbb{E} \left[\log_2 \left(1 + P_{1,2}^r |\mathbf{h}_{1,2}^* \mathbf{f}_{2,IC}|^2 + P_{1,3}^r |\mathbf{h}_{1,3}^* \mathbf{f}_{3,IC}|^2 \right) \right] \\ &\overset{(a)}{\leq} \log_2 \left(1 + P_{1,2}^r \mathbb{E} \left[|\mathbf{h}_{1,2}^* \mathbf{f}_{2,IC}|^2 \right] + P_{1,3}^r \mathbb{E} \left[|\mathbf{h}_{1,3}^* \mathbf{f}_{3,IC}|^2 \right] \right) \\ &\overset{(b)}{=} \log_2 \left(1 + P_{1,2}^r 2^{-\frac{B_I}{N_t - 1}} + P_{1,3}^r 2^{-\frac{B_I}{N_t - 1}} \right) \\ &\overset{(c)}{\leq} \log_2 \left(1 + 2P_0 \cdot 2^{-\frac{B_I}{N_t - 1}} \right), \end{aligned} \tag{5.30}$$

where step (a) follows Jensen's inequality, step (b) is from *Lemma 12*, and step (c) is due to the fact $P_{1,2}^r \leq P_0$ and $P_{1,3}^r \leq P_0$. Then by solving

$$\log_2 \left(1 + 2P_0 \cdot 2^{-\frac{B_I}{N_t - 1}} \right) = \log_2 \delta_R \tag{5.31}$$

we get the result in (5.23).

Chapter 6

Conclusion

6.1 Summary

In this dissertation, we have investigated the impact of self-interference in cellular MIMO networks. Although theoretically promising, the application of MIMO communication faces great challenges in cellular networks. Imperfect CSIT, due to such effects as feedback delay and channel quantization, causes residual inter-user interference for MU-MIMO systems, which limits the performance at high SNR; other-cell interference degrades the performance of MIMO in cellular networks, and BS coordination, which provides significant throughput gain and eliminates other-cell interference under ideal assumptions, is of very high complexity and overhead. In this dissertation, we have proposed adaptive strategies with practical transmission techniques to deal with both inter-user interference and other-cell interference.

With imperfect CSIT, the throughput of MU-MIMO systems is limited by residual inter-user interference at high SNR. We analytically compared the achievable throughput of both SU-MIMO and MU-MIMO considering feedback delay and channel quantization. It was shown that MU-MIMO can outperform SU-MIMO at medium SNR if the delay is small and the number of feedback bits is large. To exploit the benefits of MU-MIMO, we proposed an adaptive mode switching algorithm that switches between the SU and MU modes to improve the throughput. As the number of active users is closely related to array gain, spatial multiplexing gain, and residual interference, further performance improvement was achieved by varying the number of active users to balance these effects, which is called *multi-mode transmission*. Furthermore, the idea of multi-mode transmission was extended for user scheduling in both homogeneous and heterogeneous networks. Related results have appeared in

[164, 165, 168, 169].

To make BS coordination practical, we proposed two different coordination strategies. The first strategy, *clustered linear precoding*, performs localized coordination to reduce the complexity, while other-cell interference is efficiently suppressed by full intra-cluster coordination and limited inter-cluster coordination. We designed the multicell precoder to perform the coordination and discussed the design of some important parameters. The second strategy, called *adaptive intercell interference cancellation*, suppresses other-cell interference by zero-forcing precoding with multiple BS antennas, while serves each user with a single BS, i.e. it is a type of single-BS coordination transmission. The transmission techniques at multiple BSs, either single-cell beamforming or intercell interference cancellation, are jointly selected to maximize the sum throughput, which only depends on user locations. Related results have appeared in [163, 166, 167].

The results in this dissertation are mainly from a collaborative project between AT&T Labs and WNCG of UT Austin. The part on residual inter-user interference for MU-MIMO systems was inspired during my summer intern at AT&T labs. Our results shed some light on how to exploit the capacity gain of MIMO communication in cellular networks, and provide design guidelines for MIMO transmission in broadband wireless networks such as WiMAX and 3GPP LTE.

- The study on MU-MIMO with imperfect CSIT shows that we should explicitly consider the level of residual interference when deploying MU-MIMO. Generally, we should serve fewer users than the maximum number supported in the system. In practice, we can divide different scenarios into several categories, and select the preferred mode for each category by simulation or measurement. Based on this, each user is able to determine its preferred transmission mode and informs the BS about its decision. The BS takes such information into consideration when doing scheduling. In this way, the active mode of each user is relatively static and does not require frequent update.

- The results on BS coordination demonstrate the effectiveness of the clustered approach, which is a practical way to implement BS coordination. The idea of grouping users into edge and interior users can be used in practical systems, where the grouping can be based on measurement. The results in Chapter 5 show that we can get a significant throughput gain even with coordinated single-cell transmission, which is highly desirable, as it has lower complexity, lower backhaul capacity requirement, and lower CSI requirement. Therefore, coordinated single-cell transmission can be deployed as the first step to exploit the performance gain of BS coordination.

Based on my PhD research and intern experience at AT&T Labs, I also co-authored a book on 3GPP LTE standards [55], which demonstrates the application of MIMO techniques in LTE and the practical system constraints we need to consider. Besides the results presented in this dissertation, I have also done some research on the distributed antenna system [161, 162], which is one way to improve the wireless coverage and reduce interference, and is also related to multi-cell processing.

6.2 Future Work

The results in this dissertation demonstrated the significant role of self-interference in cellular MIMO networks. There are some interesting open issues and topics for future research, and this final section will discuss a few of them that are closely related to this dissertation.

First, the study in Chapter 2 and Chapter 3 assumed CSIT of fixed accuracy, i.e. with a fixed number of feedback bits, and a random quantization codebook. The accuracy of the available CSIT can be improved by exploiting the channel statistics and carefully designing the codebook and the feedback strategy. This aspect has drawn lots of attention recently and encouraging results have appeared, such as the one on predictive vector quantization [70] and another on feedback with progressive refinement [150]. Further investigation is required on how to further improve the CSIT

accuracy and how to implement such techniques in practical systems. In addition, the error in the feedback channel should also be considered.

Although limited feedback has been explored extensively in the single-cell scenario to provide partial CSIT, there is little investigation on how to do feedback efficiently in BS coordination systems. The cluster based coordination in Chapter 4 can reduce the amount of feedback, but the feedback overhead is still daunting compared to conventional single-cell processing. Chapter 5 made a brief discussion on the feedback design in the adaptive inter-cell interference cancellation system, which provided some insights and may be extended to more general scenarios. Besides instantaneous CSIT, channel statistics information such as mean and variance, and channel correlation information can also be exploited to suppress other-cell interference throughput BS coordination. Such information changes slowly with time and is easy to get.

For the MU-MIMO transmission in this dissertation, either in the single-cell or the multicell scenario, only the linear precoder with zero-forcing criterion was considered. How to design precoders that are robust to imperfect CSIT requires further investigation. In particular, it is not clear what is the optimal precoder with imperfect CSIT. For the precoder in the inter-cell interference cancellation system in Chapter 5, only selfish beamforming that maximizes the received signal power and zero-forcing precoding that minimizes the interference power are considered, and the precoder that balances these two effects can be developed to improve the performance. There is another interesting direction to improve the performance of MU-MIMO with imperfect CSIT: explicitly consider the residual interference and treat the system as an interference channel [26, 44, 121], so the techniques developed for the interference channel may be applied to improve the performance, such as the Han-Kobayashi scheme [59].

Bibliography

[1] 3GPP TR 25.913 V8.0.0. Requirements for Evolved UTRA (E-UTRA) and Evolved UTRAN (E-UTRAN) (Release 8). Dec. 2008.

[2] 3GPP TR 36.913 V8.0.1. Requirements for further advancements for Evolved Universal Terrestrial Radio Access (E-UTRA) (LTE-Advanced) (Release 8). Mar. 2009.

[3] 3GPP TS 36.211 V8.5.0. Evolved Universal Terrestrial Radio Access (E-UTRA); Physical Channels and Modulation (Release 8). Dec. 2008.

[4] 3GPP TSG RAN WG1 R1-050507. Soft frequency reuse scheme for UTRAN LTE. Huawei, Tech. Rep., Athens, Greece, May 2005.

[5] 3GPP TSG RAN WG1 R1-050764. Inter-cell interference handling for E-UTRA. Ericsson, Tech. Rep., 2005.

[6] 3GPP TSG RAN WG1 R1-060135. Interference mitigation by partial frequency reuse. Siemens, Tech. Rep., Helsinki, Finland, Jan. 2006.

[7] D. Aktas, M. N. Bacha, J. S. Evans, and S. V. Hanly. Scaling results on the sum capacity of cellular networks with MIMO links. *IEEE Trans. Inform. Theory*, 52(7):3264–3274, Jul. 2006.

[8] E. Aktas, J. Evans, and S. Hanly. Distributed decoding in a cellular multiple access channel. *IEEE Trans. Wireless Commun.*, 7(1):241–250, Jan. 2008.

[9] S. Alamouti. A simple transmit diversity technique for wireless communications. *IEEE J. Select. Areas Commun.*, 16(8):1451–1458, Oct. 1998.

[10] M. Alouini and A. Goldsmith. Capacity of Rayleigh fading channels under different adaptive transmission and diversity-combining techniques. *IEEE Trans. Veh. Technol.*, 48(4):1165–1181, Jul. 1999.

[11] J. Andrews, A. Ghosh, and R. Muhamed. *Fundamentals of WiMAX*. Prentice Hall, 2007.

[12] J. G. Andrews, W. Choi, and R. W. Heath Jr. Overcoming interference in spatial multiplexing MIMO cellular networks. *IEEE Wireless Communications Magazine*, 14(6):95–104, Dec. 2007.

[13] C. K. Au-Yeung and D. J. Love. On the performance of random vector quantization limited feedback beamforming in a MISO system. *IEEE Trans. Wireless Commun.*, 6(2):458–462, Feb. 2007.

[14] C. K. Au-Yeung, S. Y. Park, and D. J. Love. A simple dual-mode limited feedback multiuser downlink system. *accepted for publication in IEEE Trans. Comm.*

[15] E. Biglieri, R. Calderbank, A. Constantinides, A. Goldsmith, A. Paulraj, and H. V. Poor. *MIMO Wireless Communications*. Cambridge University Press, Cambridge, 2007.

[16] R. S. Blum. MIMO capacity with interference. *IEEE J. Select. Areas Commun.*, 21(5):793–801, Jun. 2003.

[17] F. Boccardi and H. Huang. Zero-forcing precoding for the MIMO broadcast channel under per-antenna power constraints. In *Proc. IEEE SPAWC*, pages 1–5, Cannes, Jul. 2006.

[18] F. Boccardi and H. Huang. Limited downlink network coordination in cellular networks. In *Proc. of the IEEE Int. Symp. on Personal Indoor and Mobile Radio Comm.*, Athens, Greece, Sept. 2007.

[19] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, 2004.

[20] G. Caire. MIMO downlink joint processing and scheduling: A survey of classical and recent results. In *Proc. Workshop on Information Theory and Its Applications*, San Diego, CA, Jan. 2006.

[21] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran. Multiuser MIMO achievable rates with downlink training and channel state feedback. *Submitted to IEEE Trans. Inform. Theory*, 2007. available online at http://arxiv.org/pdf/0711.2642.

[22] G. Caire, N. Jindal, M. Kobayashi, and N. Ravindran. Quantized vs. analog feedback for the MIMO broadcast channel: A comparison between zero-forcing based achievable rates. In *Proc. IEEE Int. Symp. Information Theory*, pages 2046–2050, Nice, France, Jun. 2007.

[23] G. Caire, N. Jindal, and S. Shamai. On the required accuracy of transmitter channel state information in multiple antenna broadcast channels. In *Proc. of the IEEE Asilomar Conf. on Signals, Systems, and Computers*, pages 289–291, Pacific Grove, CA, Nov. 2007.

[24] G. Caire, S. A. Ramprashad, H. C. Papadoupoulos, C. Pepin, and C.-E. W. Sundberg. Multiuser MIMO downlink with limited inter-cell cooperation: Approximate interference alignment in time, frequency, and space. In *Proc. of the Allerton Conf. on Comm. Control and Comp.*, Monticello, IL, Sept. 2008.

[25] G. Caire and S. Shamai (Shitz). On the achievable throughput of a multiantenna Gaussian broadcast channel. *IEEE Trans. Inform. Theory*, 49(7):1691–1706, Jul. 2003.

[26] A. B. Carleial. Interference channels. *IEEE Trans. Inform. Theory*, 24(1):60–70, Jan. 1978.

[27] S. Catreux, P. F. Driessen, and L. J. Greenstein. Simulation results for an interference-limited multiple-input multiple-output cellular system. *IEEE Comm. Lett.*, 4:334–336, Nov. 2000.

[28] C. B. Chae, D. Mazzarese, T. Inoue, and R. W. Heath Jr. Coordinated beamforming for the multiuser MIMO broadcast channel with limited feedforward. *IEEE Trans. Signal Processing*, 56(12):6044–6056, Dec. 2008.

[29] C. B. Chae, D. Mazzarese, N. Jindal, and R. W. Heath Jr. Coordinated beamforming with limited feedback in the MIMO broadcast channel. *IEEE J. Select. Areas Commun.*, 26(8):1505–1515, Oct. 2008.

[30] R. Chen, J. G. Andrews, and R. W. Heath Jr. Multiuser space-time block coded MIMO system with downlink precoding. In *Proc. IEEE Int. Conf. Commun.*, volume 5, pages 2689–2693, Paris, France, Jun. 2004.

[31] R. Chen, R. W. Heath Jr., and J. G. Andrews. Transmit selection diversity for multiuser spatial multiplexing systems with linear receivers. *IEEE Trans. Signal Processing*, 55(3):1159–1171, Mar. 2007.

[32] L. U. Choi and R. D. Murch. A transmit preprocessing technique for multiuser MIMO systems using a decomposition approach. *IEEE Trans. Wireless Commun.*, 3(1):20–24, Jan. 2004.

[33] W. Choi and J. G. Andrews. The capacity gain from intercell scheduling in multi-antenna systems. *IEEE Trans. Wireless Commun.*, 7(2):714–725, Feb. 2008.

[34] C. Chuah, D. N. C. Tse, J. Kahn, and R. Valenzuela. Capacity scaling in dual-antenna-array wireless systems. *IEEE Trans. Inform. Theory*, 48(3):637–650, Mar. 2002.

[35] R. H. Clarke. A statistical theory of mobile radio reception. *Bell System Tech. J.*, 47:957–1000, 1968.

[36] M. Costa. Writing on dirty paper. *IEEE Trans. Inform. Theory*, 39(3):439–441, May 1983.

[37] H. Dahrouj and W. Yu. Coordinated beamforming for the multi-cell multi-antenna wireless system. In *Conference on Information Sciences and Systems (CISS)*, Princeton, NJ, Mar. 2008.

[38] H. Dai, A. Molisch, and H. Poor. Downlink capacity of interference-limited MIMO systems with joint detection. *IEEE Trans. Wireless Commun.*, 3(2):442–453, Mar. 2004.

[39] W. Dai, Y. Liu, B. C. Rider, and W. Gao. How many users should be turned on in a multi-antenna broadcast channel? *IEEE J. Select. Areas Commun.*, 26(8):1526–1535, Oct. 2008.

[40] S. Das, H. Viswanathan, and G. Rittenhouse. Dynamic load balancing through coordinated scheduling in packet data systems. pages 786–796, San Francisco, CA, Apr. 2003.

[41] R. de Francisco and D. T. M. Slock. A design framework for scalar feedback in MIMO broadcast channels. *EURASIP Journal on Advances in Signal Processing*, 8(1):1–12, 2008.

[42] G. Dimic and N. D. Sidiropoulos. On downlink beamforming with greedy user selection: performance analysis and a simple new algorithm. *IEEE Trans. Signal Processing*, 53(10):3857–3868, Oct. 2005.

[43] P. Ding, D. J. Love, and M. D. Zoltowski. Multiple antenna broadcast channels with partial and limited feedback. *IEEE Trans. Signal Processing*, 55(7):3417–3428, Jul. 2007.

[44] R. Etkin, D. Tse, and H. Wang. Gaussian interference channel capacity to within one bit. *IEEE Trans. Inform. Theory*, 54(12):5534–5562, Dec. 2008.

[45] F. R. Farrokhi, K. R. Liu, and L. Tassiulas. Transmit beamforming and power control for cellular wireless systems. *IEEE J. Select. Areas Commun.*, 16(8):1437–1450, Oct. 1998.

[46] G. J. Foschini and M. J. Gans. On limits of wireless communications in a fading environment when using multiple antennas. *Wireless Personal Commun.*, 6(3):311, Mar. 1998.

[47] G. J. Foschini, H. Huang, K. Karakayali, R. A. Valenzuela, and S. Venkatesan. The value of coherent base station coordination. In *Proc., Conference on Information Sciences and Systems (CISS)*, Johns Hopkins University, Mar. 2005.

[48] G. J. Foschini, K. Karakayali, and R. A. Valenzuela. Coordinating multiple antenna cellular networks to achieve enormous spectral efficiency. *IEE Proceedings*, 153(4):548–555, Aug. 2006.

[49] D. Gesbert and M. Slim Alouini. How much feedback is multi-user diversity really worth? In *Proc. IEEE Int. Conf. Commun.*, pages 234–238, Paris, France, Jun. 2004.

[50] D. Gesbert, A. Hjorungnes, and H. Skjevling. Cooperative spatial multiplexing with hybrid channel knowledge. In *IEEE International Zurich Seminar on Communications*, 2006.

[51] D. Gesbert, S. G. Kiani, A. Gjendemsjø, and G. E. Øien. Adaptation, coordination, and distributed resource allocation in interference-limited wireless networks. *Proceeding of the IEEE*, 95(12):2393–2409, Dec. 2007.

[52] D. Gesbert and M. Kountouris. Joint power control and user scheduling in multicell wireless networks: Capacity scaling laws. *submitted to IEEE Trans. on Information Theory*, Sept. 2007.

[53] D. Gesbert, M. Kountouris, R. W. Heath Jr., C. B. Chae, and T. Salzer. Shifting the MIMO paradigm: From single user to multiuser communications. *IEEE Signal Processing Magazine*, 24(5):36–46, Sept. 2007.

[54] D. Gesbert, M. Shafi, D. Shiu, P. Smith, and A. Naguib. From theory to practice: An overview of space-time coded MIMO wireless systems. *IEEE J. Select. Areas Commun.*, 21(3):281–302, Apr. 2003.

[55] A. Ghosh, J. Zhang, J. Andrews, and R. Muhamed. *Understanding LTE.* Prentice Hall, 2010. to appear.

[56] A. Goldsmith. *Wireless Communications*. Cambridge University Press, 2005.

[57] A. Goldsmith, S. A. Jafar, N. Jindal, and S. Vishwanath. Capacity limits of MIMO channels. *IEEE J. Select. Areas Commun.*, 21(5):684–702, Jun. 2003.

[58] I. S. Gradshteyn and I. M. Ryzhik. *Table of Integrals, Series, and Products*. Academic, San Diego, CA, 5 edition, 1994.

[59] T. S. Han and K. Kobayashi. A new achievable rate region for the interference channel. *IEEE Trans. Inform. Theory*, 27(1):49–60, Jan. 1981.

[60] V. Hassel, M.-S. Alouini, D. Gesbert, and G. Oien. Exploiting multiuser diversity using multiple feedback thresholds. In *Proc. IEEE Veh. Technol. Conf.*, pages 1302–1306, Stockholm, Sweden, Jun. 2005.

[61] B. Hassibi and M. Sharif. Fundamental limits in MIMO broadcast channels. *IEEE J. Select. Areas Commun.*, 25(7):1333–1344, Sept. 2007.

[62] S. Haykin. *Adaptive Filter Theory*. Prentice-Hall, Englewood Cliffs, NJ, 3 edition, 1996.

[63] M. L. Honig, U. Madhow, and S. Verdu. Blind adaptive multiuser detection. *IEEE Trans. Inform. Theory*, 41:944–960, Jul. 1995.

[64] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, U.K., 1985.

[65] H. Huang and R. A. Valenzuela. Fundamental simulated performance of downlink fixed wireless cellular networks with multiple antennas. In *Proc. of the IEEE Int. Symp. on Personal Indoor and Mobile Radio Comm.*, pages 161–165, Berlin, Germany, Sept. 2005.

[66] H. Huang and S. Venkatesan. Asymptotic downlink capacity of coordinated cellular network. In *Proc. of the IEEE Asilomar Conf. on Signals, Systems, and Computers*, pages 850–855, Pacific Grove, CA, Nov. 2004.

[67] K. Huang, J. G. Andrews, R. W. Heath Jr., D. Guo, and R. Berry. Spatial interference cancelation for mobile ad hoc networks. *IEEE Trans. Inform. Theory*, Jul. 2008. submitted.

[68] K. Huang, R. W. Heath Jr., and J. G. Andrews. Space division multiple access with a sum feedback rate constraint. *IEEE Trans. Signal Processing*, 55(7):3879–3891, Jul. 2007.

[69] K. Huang, R. W. Heath Jr., and J. G. Andrews. Limited feedback beamforming over temporally correlated channels. *IEEE Trans. Signal Processing*, 57(5):1959–1975, May 2009.

[70] T. Inoue and R. W. Heath Jr. Predictive coding on the Grassmann manifold. *IEEE Trans. Signal Processing*, Aug. 2009. submitted.

[71] International Telecommunication Union. Measuring the information society: The ICT development index. pages 1–108, 2009.

[72] S. A. Jafar and M. Fakhereddin. Degrees of freedom for the MIMO interference channel. *IEEE Trans. Inform. Theory*, 53(7):2637–2642, Jul. 2007.

[73] S. A. Jafar, G. Foschini, and A. J. Goldsmith. Phantomnet: Exploring optimal multicellular multiple antenna systems. *EURASIP Journal on Appl. Signal Processing, Special issue on MIMO Comm. and Signal Processing*, pages 591–605, May 2004.

[74] N. Jindal. A high SNR analysis of MIMO broadcast channels. In *Proc. IEEE Int. Symp. Information Theory*, pages 2310–2314, Adelaide, Australia, Sept. 2005.

[75] N. Jindal. MIMO broadcast channels with finite rate feedback. *IEEE Trans. Inform. Theory*, 52(11):5045–5059, Nov. 2006.

[76] N. Jindal, J. G. Andrews, and S. Weber. Rethinking MIMO for wireless networks: Linear throughput increases with multiple receive antennas. In *Proc. IEEE Int. Conf. Commun.*, pages 1–6, Dresden, Germany, Jun. 2009.

[77] N. Jindal and A. Goldsmith. Dirty paper coding versus TDMA for MIMO broadcast channel. *IEEE Trans. Inform. Theory*, 51(5):1783–1794, May 2005.

[78] N. Jindal, W. Rhee, S. Vishwanath, S. A. Jafar, and A. Goldsmith. Sum power iterative water-filling for multi-antenna Gaussian broadcast channels. *IEEE Trans. Inform. Theory*, 51(4):1570–1580, Apr. 2005.

[79] S. Jing, D. N. C. Tse, J. Hou, J. B. Soriaga, J. E. Smee, and R. Padovani. Multi-cell downlink capacity with coordinated processing. *EURASIP Journal on Wireless Communications and Networking*, 2008. Volume 2008, Article ID 586878.

[80] E. A. Jorswieck, P. Svedman, and B. Ottersten. Performance of TDMA and SDMA based opportunistic beamforming. *IEEE Trans. Wireless Commun.*, 7(11):4058–4063, Nov. 2008.

[81] V. Jungnickel, T. Wirth, M. Schellmann, T. Haustein, and W. Zirwas. Synchronization of cooperative base stations. In *IEEE International Symposium on Wireless Communication Systems*, pages 329–334, Oct. 2008.

[82] A. Kansal, S. N. Batalama, and D. A. Pados. Adaptive maximum SINR RAKE filtering for DS-CDMA multipath fading channels. *IEEE J. Select. Areas Commun.*, 16:1765–1773, Dec. 1998.

[83] K. Karakayali, G. J. Foschini, and R. A. Valenzuela. Network coordination for spectrally efficient communications in cellular systems. *IEEE Wireless Communications Magazine*, 13(4):56–61, Aug. 2006.

[84] K. Karakayali, G. J. Foschini, R. A. Valenzuela, and R. D. Yates. On the maximum common rate achievable in a coordinated network. In *Proc. IEEE Int. Conf. Commun.*, pages 4333–4338, Istanbul, Turkey, June 2006.

[85] K. Karakayali, R. Yates, G. Foschini, and R. Valenzuela. Optimum zero-forcing beamforming with per-antenna power constraints. In *Proc. IEEE Int. Symp. Information Theory*, pages 101–105, Nice, France, Jun. 2007.

[86] S. G. Kiani, D. Gesbert, and G. E. Øien. Maximizing multi-cell capacity using distributed power allocation and scheduling. In *Proc. IEEE Wireless Commun. Networking Conf.*, pages 1690–1694, Hong Kong, Mar. 2007.

[87] Saad G. Kiani and D. Gesbert. Optimal and distributed scheduling for multicell capacity maximization. *IEEE Trans. Wireless Commun.*, 7(1):288–297, Jan. 2008.

[88] R. Knopp and P. Humblet. Information capacity and power control in single cell multiuser communications. In *Proc. IEEE Int. Conf. Commun.*, pages 331–335, Seattle, WA, Jun. 1995.

[89] M. Kobayashi and G. Caire. Joint beamforming and scheduling for a multi-antenna downlink with imperfect transmitter channel knowledge. *IEEE J. Select. Areas Commun.*, 25(7):1468–1477, Sept. 2007.

[90] M. Kountouris, R. de Francisco, D. Gesbert, D. T. M. Slock, and T. Salzer. Efficient metrics for scheduling in MIMO broadcast channels with limited feedback. In *Proc. of the IEEE Int. Conf. on Acoustics, Speech, and Signal Proc.*, pages 109–112, Honolulu, USA, Apr. 2007.

[91] M. Kountouris, D. Gesbert, and T. Sälzer. Distributed transmit mode selection for MISO broadcast channels with limited feedback: Switching from SDMA to TDMA. In *Proc. IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC 2008)*, pages 371–375, Recife, Brazil, Jul. 2008.

[92] A. Lapidoth, S. Shamai, and M. A. Wigger. On the capacity of fading MIMO broadcast channels with imperfect transmitter side-information. In *Proceedings of the 43rd Allerton Conference on Communication, Control and Computing*, Monticello, IL, Sept. 2005.

[93] V. K. N. Lau and M. Jiang. Performance analysis of multiuser downlink space-time scheduling for TDD systems with imperfect CSIT. *IEEE Trans. Veh. Technol.*, 55(1):296–305, Jan. 2006.

[94] C. Lee, C. B. Chae, S. Vishwanath, and R. W. Heath Jr. Adaptive mode switching in correlated multiple antenna cellular networks. *Journal of Communications and Networks*, 11(3):279–286, Jun. 2009.

[95] J. Lee, J.-K. Han, and J. Zhang. MIMO technologies in 3GPP LTE and LTE-Advanced. *EURASIP Journal on Wireless Communications and Networking*, 2009. Volume 2009, Article ID 302092, 10 pages.

[96] J. Lee and N. Jindal. High SNR analysis for MIMO broadcast channels: Dirty paper coding vs. linear precoding. *IEEE Trans. Inform. Theory*, 53(12):4787–4792, Dec. 2007.

[97] D. J. Love, R. W. Heath Jr., V. K. N. Lau, D. Gesbert, B. D. Rao, and M. Andrews. An overview of limited feedback in wireless communication systems. *IEEE J. Select. Areas Commun.*, 26(8):1341–1365, Oct. 2008.

[98] D. J. Love, R. W. Heath Jr., W. Santipach, and M. Honig. What is the value of limited feedback for MIMO channels? *IEEE Commun. Mag.*, 42(10):54–59, Oct. 2004.

[99] D. J. Love, R. W. Heath Jr., and T. Strohmer. Grassmannian beamforming for multiple-input multiple-output wireless systems. *IEEE Trans. Inform. Theory*, 49(10):2735–2745, Oct. 2003.

[100] D. J. Love and R. W. Heath Jr. Multi-mode precoding using linear receivers for limited feedback MIMO systems. In *Proc. IEEE Int. Conf. Commun.*, pages 448–452, Paris, France, Jun. 2004.

[101] P. Marsch and G. Fettweis. A framework for optimizing the downlink performance of distributed antenna systems under a constrained backhaul. In *Proc. European Wireless Conf. (EW' 07)*, Paris, France, Apr. 2007.

[102] P. Marsch and G. Fettweis. A framework for optimizing the uplink performance of distributed antenna systems under a constrained backhaul. In *Proc. IEEE Int. Conf. Commun.*, pages 975–979, Glasgow, Scotland, Jun. 2007.

[103] K. Mukkavilli, A. Sabharwal, E. Erkip, and B. Aazhang. On beamforming with finite rate feedback in multiple-antenna systems. *IEEE Trans. Inform. Theory*, 49(10):2562–2579, Oct. 2003.

[104] A. Narula, M. J. Lopez, M. D. Trott, and G. W. Wornell. Efficient use of side information in multiple-antenna data transmission over fading channels. *IEEE J. Select. Areas Commun.*, 16:1423–1436, Oct. 1998.

[105] B. Ng, J. Evans, and S. Hanly. Distributed linear multiuser detection in cellular networks based on Kalman smoothing. In *Proc. IEEE Global Telecommunications Conf.*, pages 134–138, Dallas, TX, Nov. 2004.

[106] B. Ng, J. Evans, S. Hanly, and D. Aktas. Distributed downlink beamforming with cooperative base stations. *IEEE Trans. Inform. Theory*, 54(12):5491–5499, Dec. 2008.

[107] H. T. Nguyen, J. B. Anderson, and G. F. Pedersen. Capacity and performance of MIMO systems under the impact of feedback delay. In *Proc. of the IEEE Int. Symp. on Personal Indoor and Mobile Radio Comm.*, pages 53–57, Barcelona, Spain, Sept. 2004.

[108] E. N. Onggosanusi, A. Gatherer, A. G. Dabak, and S. Hosur. Performance analysis of closed-loop transmit diversity in the presence of feedback delay. *IEEE Trans. Commun.*, 49(9):1618–1630, Sept. 2001.

[109] D. A. Pados and S. N. Batalama. Joint spacetime auxiliary-vector filtering for DS/CDMA systems with antenna arrays. *IEEE Trans. Commun.*, 47:1406–1415, Sept. 1999.

[110] Z. Pan, K. K. Wong, and T. S. Ng. Generalized multiuser orthogonal space-division multiplexing. *IEEE Trans. Wireless Commun.*, 3(6):1969–1973, Nov. 2004.

[111] A. Papadogiannis, D. Gesbert, and E. Hardouin. A dynamic clustering approach in wireless networks with multi-cell cooperative processing. In *Proc. IEEE Int. Conf. Commun.*, pages 4033–4037, Beijing, China, May 2008.

[112] A. Paulraj, R. Nabar, and D. Gore. *Introduction to Space-Time Wireless Communications*. Cambridge University Press, Cambridge, 2003.

[113] C. Peel, B. Hochwald, and A. Swindlehurst. Vector-perturbation technique for near-capacity multiantenna multiuser communication-Part I: Channel inversion and regularization. *IEEE Trans. Commun.*, 53(1):195–202, Jan. 2005.

[114] N. Ravindran and N. Jindal. Limited feedback-based block diagonalization for the MIMO broadcast channel. *IEEE J. Select. Areas Commun.*, 26(8):1473–1482, Oct. 2008.

[115] N. Ravindran, N. Jindal, and H. C. Huang. Beamforming with finite rate feedback for LOS MIMO downlink channels. In *Proc. IEEE Global Telecommunications Conf.*, pages 4200–4204, Washington, DC, Nov. 2007.

[116] C. B. Ribeiro, K. Hugl, M. Lampinen, and M. Kuusela. Performance of linear multi-user MIMO precoding in LTE system. In *3rd International Symposium on Wireless Pervasive Computing*, pages 410–414, May 2008.

[117] J. C. Roh and B. D. Rao. Design and analysis of MIMO spatial multiplexing systems with quantized feedback. *IEEE Trans. Signal Processing*, 54(8):2874–2886, Aug. 2006.

[118] A. Sanderovich, O. Somekh, and S. Shamai (Shitz). Uplink macro diversity with limited backhaul capacity. In *Proc. IEEE Int. Symp. Information Theory*, pages 11–15, Nice, France, Jun. 2007.

[119] A. Sang, X. Wang, M. Madihian, and R. D. Gitlin. Coordinated load balancing, handoff/cell-site selection, and scheduling in multi-cell packet data systems. *Wireless Networks*, 14(1):103–120, Jan. 2008.

[120] W. Santipach and M. Honig. Asymptotic capacity of beamforming with limited feedback. In *Proc. IEEE Int. Symp. Information Theory*, page 290, Chicago, IL, Jun./Jul. 2004.

[121] H. Sato. The capacity of the Gaussian interference channel under strong interference. *IEEE Trans. Inform. Theory*, 27(6):786–788, Nov. 1981.

[122] S. Shamai (Shitz), O. Somekh, and B. M. Zaidel. Multi-cell communications: An information theoretic perspective. In *Joint Workshop on Communications and Coding (JWCC)*, Florence, Italy, Oct. 2004.

[123] S. Shamai (Shitz) and B. M. Zaidel. Enhancing the cellular downlink capacity via co-processing at the transmitting end. In *Proc. IEEE Veh. Technol. Conf.*, pages 1745–1749, Rhodes, Greece, May 2001.

[124] M. Sharif and B. Hassibi. On the capacity of MIMO broadcast channels with partial side information. *IEEE Trans. Inform. Theory*, 51(2):506–522, Feb. 2005.

[125] M. Sharif and B. Hassibi. A comparison of time-sharing, DPC, and bemforming for MIMO broadcast channels with many users. *IEEE Trans. Commun.*, 55(1):11–15, Jan. 2007.

[126] Z. Shen, R. Chen, J. G. Andrews, R. W. Heath Jr., and B. L. Evans. Low complexity user selection algorithms for multiuser MIMO systems with block diagonalization. *IEEE Trans. Signal Processing*, 54(9):3658–3663, Sept. 2006.

[127] Z. Shen, R. Chen, J. G. Andrews, R. W. Heath Jr., and B. L. Evans. Sum capacity of multiuser MIMO broadcast channels with block diagonalization. *IEEE Trans. Wireless Commun.*, 6(6):2040–2045, Jun. 2007.

[128] S. Shim, J. S. Kwak, R. W. Heath Jr., and J. G. Andrews. Block diagonalization for multi-user MIMO with other-cell interference. *IEEE Trans. Wireless Commun.*, 7(7):2671–2681, Jul. 2008.

[129] M. K. Simon. *Probability Distributions Involving Gaussian Random Variables: A Handbook for Engineers and Scientists*. Springer, 2002.

[130] H. Skjevling, D. Gesbert, and A. Hjorungnes. Receiver-enhanced cooperative spatial multiplexing with hybrid channel knowledge. In *Proc. of the IEEE Int. Conf. on Acoustics, Speech, and Signal Proc.*, pages 65–68, Toulouse, France, May 2006.

[131] O. Somekh, O. Simeone, Y. Bar-Ness, and A. M. Haimovich. Distributed multi-cell zero-forcing beamforming in cellular downlink channels. In *Proc. IEEE Global Telecommunications Conf.*, pages 1–6, San Francisco, Nov. 2006.

[132] O. Somekh, B. M. Zaidel, and S. Shamai (Shitz). Sum rate characterization of joint multiple cell-site processing. *IEEE Trans. Inform. Theory*, 53(12):4473–4497, Dec. 2007.

[133] Q. H. Spencer, A. L. Swindlehurst, and M. Haardt. Zero-forcing methods for downlink spatial multiplexing in multi-user MIMO channels. *IEEE Trans. Signal Processing*, 52(2):461–471, Feb. 2004.

[134] V. Tarokh, H. Jafarkhani, and A. Calderbank. Space-time block codes from orthogonal designs. *IEEE Trans. Inform. Theory*, 45(5):1456–1467, Jul. 1999.

[135] V. Tarokh, N. Seshadri, and A. R. Calderbank. Space-time codes for high data rate wireless communication: Performance analysis and code construction. *IEEE Trans. Inform. Theory*, 44(2):744–765, Mar. 1998.

[136] I. E. Telatar. Capacity of mulit-antenna Gaussian channels. *Europ. Trans. Telecommun.*, 10:585–595, Nov. 1999.

[137] M. Trivellato, F. Boccardi, and F. Tosato. User selection schemes for MIMO broadcast channels with limited feedback. In *Proc. IEEE Veh. Technol. Conf.*, Dublin, Ireland, Apr. 2007.

[138] David Tse and Pramod Viswanath. *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.

[139] A. M. Tulino, S. Verdú, and A. Lozano. Capacity of antenna arrays with space, polarization and pattern diversity. In *Proc. IEEE Information Theory Workshop*, pages 324–327, Paris, France, Mar. 2003.

[140] A. Vakili, A. F. Dana, and B. Hassibi. On the throughput of opportunistic beamforming with imperfect CSI. In *Proc. of ACM IWCMC*, pages 19–23, Honolulu, HI, Aug. 2007.

[141] A. Vakili and B. Hassibi. On the throughput of broadcast channels with imperfect CSI. In *Proc. of IEEE SPAWC*, pages 1–5, Cannes, France, Jul. 2006.

[142] A. Vakili, M. Sharif, and B. Hassibi. The effect of channel estimation error on the throughput of broadcast channels. In *Proc. of the IEEE Int. Conf. on Acoustics, Speech, and Signal Proc.*, pages IV 29–32, Toulouse, France, May 2006.

[143] S. Venkatesan. Coordinating base stations for greater uplink spectral efficiency in a cellular network. In *Proc. of the IEEE Int. Symp. on Personal Indoor and Mobile Radio Comm.*, pages 1–5, Athens, Greece, Sept. 2007.

[144] S. Venkatesan, H. C. Huang, A. Lozano, and R. Valenzuela. A WiMAX-based implementation of network MIMO for indoor wireless systems. *EURASIP Journal on Advances in Signal Processing*, 2009.

[145] S. Vishwanath, N. Jindal, and A. Goldsmith. Duality, achievable rates, and sum-rate capacity of MIMO broadcast channels. *IEEE Trans. Inform. Theory*, 49(10):2658–2668, Oct. 2003.

[146] P. Viswanath and D. Tse. Sum capacity of the vector Gaussian broadcast channel and uplink-downlink duality. *IEEE Trans. Inform. Theory*, 49(8):1912–1921, Aug. 2003.

[147] P. Viswanath, D. N. C. Tse, and R. Laroia. Opportunistic beamforming using dumb antennas. *IEEE Trans. Inform. Theory*, 48(6):1277–1294, Jun. 2002.

[148] R. W. Heath Jr. and David J. Love. Multimode antenna selection for spatial multiplexing systems with linear receivers. *IEEE Trans. Signal Processing*, 53(8):3042–3056, Aug. 2005.

[149] R. W. Heath Jr. and A. J. Paulraj. Switching between diversity and multiplexing in MIMO systems. *IEEE Trans. Commun.*, 53(6):962–968, Jun. 2005.

[150] R. W. Heath Jr., T. Wu, and A. C. K. Soong. Progressive refinement for high resolution limited feedback beamforming. *EURASIP Journal on Advances in Signal Processing*, 2009. Article ID 463823.

[151] H. Weingarten, Y. Steinberg, and S. Shamai. The capacity region of the Gaussian multiple-input multiple-output broadcast channel. *IEEE Trans. Inform. Theory*, 52(9):3936–3964, Sept. 2006.

[152] J. H. Winters. Optimum combining in digital mobile radio with cochannel interference. *IEEE Trans. Veh. Technol.*, 33(3):523–539, Aug. 1984.

[153] J. H. Winters, J. Salz, and R. D. Gitlin. The impact of antenna diversity on the capacity of wireless communication systems. *IEEE Trans. Commun.*, 42(2,3,4):1740–1751, Feb/Mar/Apr 1994.

[154] T. Wu and V. K. N. Lau. Robust rate, power and precoder adaptation for slow fading MIMO channels with noisy limited feedback. *IEEE Trans. Wireless Commun.*, 7(6):2360–2367, Jun. 2008.

[155] Y. Xiao. IEEE 802.11n: enhancements for higher throughput in wireless LANs. *IEEE Wireless Communications*, 12(6):82–91, Dec. 2005.

[156] T. Yoo, N. Jindal, and A. Goldsmith. Multi-antenna downlink channels with limited feedback and user selection. *IEEE J. Select. Areas Commun.*, 25(7):1478–1491, Sept. 2007.

[157] W. Yu and J. Cioffi. The sum capacity of a Gaussian vector broadcast channel. *IEEE Trans. Inform. Theory*, 50(9):1875–1892, Sep. 2004.

[158] W. Yu and T. Lan. Transmitter optimization for the multi-antenna downlink with per-antenna power constraints. *IEEE Trans. Signal Processing*, 55(6):2646–2660, Jun. 2007.

[159] H. Zhang and H. Dai. Cochannel interference mitigation and cooperative processing in downlink multicell multiuser MIMO networks. *EURASIP Journal on Wireless Communications and Networking*, (2):222–235, 4th Quarter 2004.

[160] H. Zhang, N. B. Mehta, A. F. Molisch, J. Zhang, and H. Dai. Asynchronous interference mitigation in cooperative base station systems. *IEEE Trans. Wireless Commun.*, 7(1):155–165, Jan. 2008.

[161] J. Zhang and J. G. Andrews. Cellular communication with randomly placed distributed antennas. In *Proc. IEEE Global Telecommunications Conf.*, pages 1400–1404, Washington, D.C., Dec. 2007.

[162] J. Zhang and J. G. Andrews. Distributed antenna systems with randomness. *IEEE Trans. Wireless Commun.*, 7(9):3636–3646, Sept. 2008.

[163] J. Zhang and J. G. Andrews. Adaptive spatial intercell interference cancellation in multicell wireless networks. *IEEE J. Select. Areas Commun.*, Sept. 2009. submitted.

[164] J. Zhang, J. G. Andrews, and R. W. Heath Jr. Single-user MIMO vs. multiuser MIMO in the broadcast channel with CSIT constraints. In *Allerton Conf. on Comm. Control and Comp.*, pages 309–314, Monticello, IL, Sept. 2008.

[165] J. Zhang, J. G. Andrews, and R. W. Heath Jr. Block diagonalization in the MIMO broadcast channel with delayed CSIT. In *Proc. IEEE Global Telecommunications Conf.*, Honolulu, USA, Nov. 2009.

[166] J. Zhang, R. Chen, J. G. Andrews, A. Ghosh, and R. W. Heath Jr. Networked MIMO with clustered linear precoding. *IEEE Trans. Wireless Commun.*, 8(4):1910–1921, Apr. 2009.

[167] J. Zhang, R. Chen, J. G. Andrews, and R. W. Heath Jr. Coordinated multi-cell MIMO systems with cellular block diagonalization. In *Proc. of the IEEE Asilomar Conf. on Signals, Systems, and Computers*, pages 1669–1673, Pacific Grove, CA, Nov. 2007.

[168] J. Zhang, M. Kountouris, J. G. Andrews, and R. W. Heath Jr. Achievable throughput of multi-mode multiuser MIMO with imperfect CSI constraints. In *Proc. IEEE Int. Symp. Information Theory*, pages 2659–2663, Seoul, Korea, Jun. 2009.

[169] J. Zhang, R. W. Heath Jr., M. Kountouris, and J. G. Andrews. Mode switching for the multi-antenna broadcast channel based on delay and channel quantization. *EURASIP Journal on Advances in Signal Processing*, 2009. Article ID 802548, 14 pages.

[170] L. Zheng and D. Tse. Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel. *IEEE Trans. Inform. Theory*, 48(2):359–383, Feb. 2002.

[171] L. Zheng and D. Tse. Diversity and multiplexing: a fundamental tradeoff in multiple-antenna channels. *IEEE Trans. Inform. Theory*, 49(5):1073–1096, May 2003.

[172] S. Zhou, Z. Wang, and G. B. Giannakis. Quantifying the power loss when transmit beamforming relies on finite rate feedback. *IEEE Trans. Wireless Commun.*, 4(7):1948–1957, Jul. 2005.

Vita

Jun Zhang is a Ph.D. candidate in the Department of Electrical and Computer Engineering at The University of Texas at Austin. He received the B.Eng. degree in Electronic Engineering from The University of Science and Technology of China (USTC), Hefei, China, in July 2004, and received his M.Phil. degree in Information Engineering from The Chinese University of Hong Kong (CUHK), Hong Kong, in August 2006. His PhD dissertation focuses on cellular MIMO networks, with the objective to efficiently suppress self-interference in cellular MIMO networks and improve the system throughput. He was an intern at AT&T Labs Inc. in the summer of 2007 and 2008, working on the application of multiuser MIMO in WiMAX and 3GPP LTE. His research interest focuses on the application of MIMO communications in cellular and ad hoc networks. He is co-author of the Prentice-Hall book, *Understanding LTE*. His PhD research has been primarily supported by AT&T Labs.

Permanent address: 3543 Greystone Dr.
Austin, Texas 78731

This dissertation was typeset with LaTeX[†] by the author.

[†]LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.